2012

Developing A Geospatial Protocol For Coral Epizootiology

Jennifer Anne Lentz
Louisiana State University and Agricultural and Mechanical College, jennifer.lentz@gmail.com

DEVELOPING A GEOSPATIAL PROTOCOL FOR CORAL EPIZOOTIOLOGY

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Oceanography and Coastal Sciences

by
Jennifer Anne Lentz
B.A., Hamilton College, 2005
May, 2012

“If I can’t picture it, I can’t understand it.”

(Albert Einstein)

“Studying spatial structures is both a requirement

for ecologists who deal with spatial data and a challenge.”

(Legendre 1993)

Acknowledgements

I would like to thank my Advisor, Dr. Nan D. Walker, the chair of my minor, Dr. Michael Leitner, and the rest of my doctoral committee (Dr. Dubravko Justic, Dr. Michael Hellberg, Dr. Richard B. Aronson, and Dr. John P. Wefel) for their help and guidance throughout this long journey. I wish to thank Philippe Mayor for providing me with the *Acropora palmata* and white-band disease (WBD) data. This was the highest resolution and most detailed dataset I was able to find, and served as the basis of the majority of the spatial analyses performed in this dissertation. I would also like to acknowledge the National Park Service and everyone involved in Mayor et al.'s (2006) study. This dissertation has benefitted from all of your involvement, and I am very grateful.

This academic milestone would not have been possible without the assistance, insight, support, and patience of Dr. Andrew J. Curtis and Dr. Jason K. Blackburn. Thank you both for watching out for me from your respective coasts, and for guiding me through the process of getting my first manuscript published. I look forward to working with you both on publishing additional material from this dissertation and future collaborations.

I would also like to thank all of the wonderful teachers I have been so fortunate to have had throughout my life-long academic journey. Specifically, I wish to thank the following teachers for the profound influence they have had on my life. Ms. Force, my sixth grade teacher, who taught me to be confident in both myself and my convictions, and fostered my love of learning. Dr. Cynthia Domack who mentored me throughout my four years at Hamilton College. and Dr. Ernest H. Williams, Jr. who not only served as my thesis advisor during my senior year, but also saw potential in me during my first two years of college and fought to help get my interdisciplinary concentration in environmental sciences

approved. Dr. Christine K. Schelten from the School for Field Studies, who served as the advisor on my directed research project during my semester abroad, and helped develop my passion for studying both *Acropora* and coral diseases. Dr. Paul LaRock and Dr. Brian Fry from LSU's Department of Oceanography and Coastal Sciences – thank you both for all the good advice you have given me and for our many good conversations. To my high school science teacher, Dr. John Menke: it is hard to put into words what a profound impact you have had (and continue to have) on my life. Both my self-designed undergraduate concentration in environmental studies and my doctoral marine research were inspired by the passion you instilled in your students for the environment.

Last, but not least, I would never have made it this far if not for the unconditional love and support so generously offered by my family and friends.

Table of Contents

List of Tables

List of Figures

Abstract

This dissertation explores how geographic information systems (GIS) and spatial statistics, specifically the techniques used to map, detect, and spatially analyze disease epidemics, could be used to advance our understanding of coral reef health. Given that different types of spatial analysis, as well as different parameter settings within each analysis, can produce noticeably different results, poor selection or improper use of a given technique would likely lead to inaccurate representations of the spatial distribution and false interpretations of the disease. For this reason, I performed a comprehensive review of the following types of exploratory spatial data analysis (ESDA): mapping and visualization methods; centrographic and distance-based point pattern analyses; spatial kernel density estimates (KDE) using single and dual versions of adaptive and fixed-distance KDEs in which the fixed-distance KDEs were performed using bandwidths calculated using 12 different estimation methods; *SaTScan*'s spatial scan statistic using both the Bernoulli and Poisson probability models; and last, local and global versions of the Moran's I and Getis-ord G spatial autocorrelation statistics. Each technique was applied to an artificial dataset with known cluster locations in order to determine which methods provided the most accurate results. These results were then used to develop different geospatial analytical protocols based on the types of coral data available, noting that the most meaningful results would be produced using local spatial statistics to analyze data of diseased colonies and colonies from the underlying coral population at risk. Last, I applied the techniques from one of the protocols to data from a 2004 White-Band Disease (WBD) outbreak on a population of *Acropora palmata* corals in the US Virgin Islands. The results of this work represent the first application of geospatial analytical techniques in visualizing the spatial nature of a coral disease and provides important information about the epizootiology of this particular outbreak. Specifically, the results indicated that WBD prevalence was low with numerous significant

disease clusters occurring throughout the study area, suggesting WBD may be caused by a ubiquitous stressor. The material presented in this dissertation will provide researchers with the necessary tools and information needed to perform the most accurate geospatial analysis possible based on the coral data available.

Chapter 1. Introduction

"Coral reefs are widely considered to be particularly vulnerable to changes in ocean temperatures, yet we understand little about the broad-scale spatio-temporal patterns that may cause coral mortality from bleaching and disease." – Selig et al. (2010)

1.1 Approach and Rationale

Over the last few decades, there has been a substantial decline in the health of coral reefs worldwide. In order to attempt to stabilize our reefs and restore some of what has been lost, we must first understand those factors contributing to or directly causing this decline. Given the highly complex nature of the problem, stringent multi-disciplinary analytical techniques are needed to better understand the spatial nature of the decline. This dissertation will attempt to advance our understanding of coral reef health by applying geospatial techniques to the problem of coral disease, specifically those used to map and spatially analyze epidemics and general public health concerns.

For the purposes of this dissertation "disease" will be defined as any impairment of an organism's vital functions, systems, organs, or cells (Stedman 2006). In contrast, "health" will be defined as an individual's ability to resist or adapt to various stresses, whether they are physical, chemical, biological, social, etc. (Meade and Earickson 2000). Thus, a coral reef is said to be diseased when the animal is no longer able to withstand or adapt to an environmental insult causing the coral's function to decline such that their ability to survive is in jeopardy. In contrast, a healthy reef would be one which was able to successfully adapt to the insult and whose survival is not at risk.

1.2 Objectives and Hypotheses to be Tested

The purpose of my doctoral research is to determine whether geospatial analytical techniques can be used to advance our understanding of coral epizootiology, and if so, how? Specifically, I will be testing the following hypotheses:

H_1: There is spatial variation in the distribution of coral diseases and geospatial analysis of this variation can advance our understanding of coral epizootiology.

H_2: Geospatial analysis can be used to detect, quantify, statistically analyze, and visualize the spatial nature of different coral diseases.

H_3: The extent of spatial dependence (the clustering or dispersion of corals) estimated by the geospatial analysis will be influenced by: (1) changes in the spatial resolution of the data, (2) changes in the spatial parameters used during the analysis, and, (3) the type of geospatial analysis being performed.

- $H_{3.1}$: Localized clusters will be more readily detectable in high-resolution data when compared to the same analysis performed on lower resolution data.
- $H_{3.2}$: The spatial parameter settings used by a given type of spatial analysis will influence the accuracy of the estimated spatial distribution of the disease.

H_4: The results of different types of geospatial analysis performed on both coral disease data and coral population data can be used to test additional hypotheses regarding coral epizootiology, such as:

- $H_{4.1}$: Diseased corals tend to be spatially clustered in areas in which the density of the underlying population is high.
- $H_{4.2}$: Diseased corals tend to be clustered in areas in which there is strong spatial aggregation (clustering) between all individuals in the underlying population.

$H_{4.3}$: Diseased corals will have a more clustered spatial distribution than non-diseased corals.

$H_{4.4}$: The areas with the highest prevalence rates will also be found to have statistically significant prevalence.

The findings of all of the above hypotheses have the potential to be quite important as they would further our understanding of the etiology of the different coral diseases. The results might also facilitate and guide microbial analyses by showing where targeted microbial testing should be performed. Last, the results of these findings have the potential to help managers make decisions on how to protect the reefs by providing geographic information on where the reefs are the most stressed and thus at a higher risk of becoming diseased.

1.3 Synopsis of Chapters

I began with a thorough literature review of the following three subjects: geospatial analysis; epidemiology; and diseases in the marine environment (Chapter 2). This is followed by an in depth review of coral reefs, their declining health, and the types of research techniques currently being used to study coral diseases (Chapter 3). I then provide a brief overview of the datasets and general methodology used throughout the remainder of the dissertation (Chapter 4). In Chapter 5, I give a detailed explanation of the types of exploratory spatial data analysis (ESDA) that were performed; as well as, the specific methods used to perform each analysis. All of the ESDAs mentioned in Chapter 5 were performed on an artificial dataset with known cluster locations. I designed the artificial dataset to have the same spatial resolution and number of diseased and non-diseased points of a real coral disease dataset. The use of an artificial dataset enabled me to not only better assess the accuracy of the cluster detection techniques used by each of the different types of geospatial analysis; but also, study how spatial scale influenced the results of various types of analysis. Additionally, I was able to use the artificial dataset to

calibrate the different types of analysis so that they would perform optimally on the real coral disease dataset. The results of the ESDA of the artificial dataset are given in Chapter 6. The strengths and weaknesses of each spatial program were assessed by comparing the results from these analyses to the locations of the pre-defined clusters in the artificial dataset. Chapter 7 concludes by laying out specific analytic protocols for the geospatial analysis of different types of coral disease data. One of the recommended protocols was then used in Chapter 8 to examine real data from a 2004 outbreak of white-band disease (WBD) in the US Virgin Islands. The material in Chapter 8 was published earlier this year as a report by Lentz et al. (2011) in *PLoS ONE*. Finally, I close with a summary of the conclusions, implications, and recommendations for future research (Chapter 9).

The material presented in Chapters 4 through 7 is currently being compiled to form a series of methods papers designed to serve as a guide for how to use geospatial analysis to study coral disease, as well as other types of marine diseases. I plan to have these manuscripts submitted to *Geospatial Health* later this year.

Chapter 2. Geospatial Analysis, Epidemiology, and Marine Diseases

> "While the spatial nature of disease processes is widely acknowledged and some epidemic models include representations of space (Cliff 1995; Holmes 1997), Bolker et al. (1995) emphasize in their review of spatial dynamics of infectious diseases that spatial statistics are underutilized in epidemiology. One promising area is the use of spatial statistics to link transmission biology with observed spatial patterns of disease (e.g., Real et al. 1992)." – Jolles et al. (2002)

2.1 Geospatial Analysis

2.1.1 A Brief Introduction to GIS and Spatial Analysis

Geographic Information Systems (GIS) were developed starting in the early 1960s as computer-based applications for processing mapped data (Lo and Yeung 2007). Today, GIS can be used as a cartographic tool to map data, as well as an analytical tool to visually identify or test for spatial patterns within the data (Curtis and Leitner 2006). GIS allows multiple datasets and types of spatial data from a specific geographic region to be displayed simultaneously as individual layers. Further the ability of GIS to link spatial data (i.e. latitude and longitude of a diseased coral) to attribute data (i.e. name of the coral disease, species of coral, etc.) greatly enhances the analytical power of the software, far beyond that of traditional statistics. As technology continues to advance, GIS technology found widespread application, making GIS one of the fastest growing computer industries (Lo and Yeung 2007).

Perhaps some of the most important advances in GIS technology have been those relating to spatial analysis. GIS provides an environment in which a spatial problem can be mapped. Insight into this problem may then be attained by mapping other data layers in order to see if there is a visible correlation between the problem and the surrounding environment. Spatial analyses provide the researcher with the ability to test these correlations, enabling the researcher to identify relationships and have statistical evidence to support the relationships. Longley et al. (2005) described Spatial Analysis as:

> …the crux of GIS because it includes all the transformations, manipulations, and methods that can be applied to geographic data to add value to them, to support decisions, and to reveal patterns and anomalies that are not immediately obvious – in other words, spatial analysis is the process by which we turn raw data into useful information, in pursuit of scientific discovery, or more effective decision making.
>
> Spatial analysis can reveal things that might otherwise be invisible – it can make what is implicit explicit.

2.1.2 Medical Geography and Spatial Epidemiology

As the name implies, the field of Medical Geography combines medical and spatial disciplines by using geographical concepts and techniques to study health-related issues (Meade and Earickson 2000; MedicineNet 2004). The idea is that by mapping the spatial attributes of a disease, information about the disease may be revealed (Meade and Earickson 2000; Koch 2005; Gao et al. 2008).

The concept of diseases being influenced through our own interactions with the surrounding environment date back more than 2,000 years to the teachings of Hippocrates, a Greek physician, who is often referred to as the "Father of medicine" (Meade and Earickson 2000; Koch 2005). More recently, the concept was used to map outbreaks and the progression of plague around Bari, Italy in 1694 (Koch 2005; Gao et al. 2008). In 1792, the exact phrase "Medical Geography" appeared in the following title of a three-volume work by Leonhard Ludwig Finke: "Notes on General Practical Medical Geography…Dealing with the History of Medical Science and Pharmacology of the Indigenous Population of the Varying States of Germany" (Koch 2005).

Perhaps the most well known application of medical geography was by Dr. John Snow (1813-1858), a London physician during the mid-late 1800s. Dr. Snow used a map of the locations of cholera deaths in relation to London streets and water pumps to visually defend his

theory that cholera was being transmitted through contaminated drinking water (Waller and Gotway 2004; Koch 2005; Elliott et al. 2006). According to legend, shortly after Dr. Snow had the handle to the Broad Street pump removed the cholera outbreak in the surrounding area sharply declined (Koch 2005; Meade and Earickson 2005). It is important to note that while Dr. Snow's cholera work did result in his designation as the "father of modern epidemiology," his Broad Street Pump conclusions were criticized both during his time and today (McLeod 2000; Waller and Gotway 2004; Koch 2005; Koch and Denike 2009). While a few of Dr. Snow's contemporaries agreed with his hypothesis, such as Dr. Robert Webb Watkins, the majority of the medical community "vilified" the "water-borne hypothesis" (Watkins 2011). The water-borne hypothesis was not accepted by medical community until after Dr. Robert Koch had identified the water-borne pathogen that caused cholera *Vibrio cholerae,* in 1883, which was 29 years after Dr. Snow had published his map and removed the handle from the Broad Street pump (Snow 2002; Koch and Denike 2009; Watkins 2011). Dr. Snow's "germ theory" was not embraced by the medical community until it had been confirmed through advances in microbiology, and it was not until the 1930s-1950s (80-100 years after Snow's death) that his work became "unequivocally" a "classic" in the epidemiological literature (Vandenbroucke 2001; Koch 2008; Snow 2008). Much of the modern criticism of Snow's work has to do with the primitiveness of his catographic techniques; however, not only was Dr. Snow was not trained as a cartographer, but his work was more than 100 years before the development of rigorous spatial statistics, medical symbolization and mapping aesthetics (McLeod 2000; Koch 2005).

While, the field of medical geography continued to slowly grow and evolve into the 1900's, it was not until the invention of computers and Geographical Information Systems (GIS) in the 1970s and early 80s that the field began to take off. In 1984, the First International Medical Geography Symposium (IMGS) was held in Nottingham, England (Joseph 1985).

Today medical geography is being used to better understand an increasingly wide variety of issues (Earickson 2007; Griffith and Christakos 2007), ranging from the global spread of AIDS (Koch 2005; Gao et al. 2008), to the changing spatial dynamics of the polio virus between 1910 and 1971 (Trevelyan et al. 2005), to the real time interactive modeling of the SARS epidemic in China (Boulos 2004) and the rabies virus in the United States (Blanton et al. 2006). The applications of geography-based medical studies appear to be limitless, especially now that technological advances have made it possible to not just map a disease surface, but to critically analyze and test for spatial relationships as well.

"Spatial epidemiology" is a specialized form of Medical Geography that incorporates more rigorous spatial, and statistical analyses, to better facilitate epidemiological studies. The creation and widespread use of GIS, coupled with modern advances in computer technology and statistical methodology, have enabled the development of this sub-discipline, by both facilitating the availability of geographically indexed health data and creating an environment in which to display and analyze this data (Gatrell et al. 1996). The purpose of spatial epidemiology is to first describe variations in the spatial distributions of diseases, and second to perform analyses on this data, the results of which will hopefully further our understanding of the disease (Ostfeld et al. 2005; Elliott et al. 2006; Lawson 2006).

Geospatial technologies are being increasingly utilized in spatial epidemiological investigations (Croner et al. 1996; Curtis 1999; Morrison et al. 2004; Lentz et al. 2011). These technologies include methods of data collection, such as: (1) global positioning system (GPS) receivers (Dwolatzky et al. 2006); (2) software designed to manipulate, analyze, and visualize large spatial datasets, the most common of which are geographic information system (GIS) and remote sensing (RS) approaches (Clarke et al. 1996; Beck et al. 2000); and last, (3) internet-based portals for data collection, display, and distribution (Mills et al. 2008). Briefly stated,

these geospatial technologies follow one of two investigative strategies: the identification of either spatial patterns or spatial associations.

When faced with a typical disease dataset, for example, morbidity or mortality cases, a variety of techniques are applied to either the individual level or aggregated data to reveal areas of positive spatial autocorrelation, more commonly referred to as "hotspots." Hotspots are areas of high intensity, usually suggesting an elevated presence of disease. The most revealing hotspot analyses use both numerator (cases of disease) and denominator (all possible individuals) populations, so that an elevated rate can be identified, which is not dependent on large population numbers (Bithell 1990; Levine and Associates 2004; Waller and Gotway 2004; Smith and Bruce 2008; Carlos et al. 2010; Cai et al. 2011). This concept is explained in greater detail when dual kernel density estimates (KDEs) are introduced in section **5.4.2.2** on page 103.

2.2 Epidemiology

2.2.1 Brief Overview of Important Disease Terminology

Before any analysis of coral diseases can be performed, the term "disease" must be defined and clarified, as the field of coral pathology is fraught with confusion, misdiagnoses, and multiple names for the same syndrome. *Disease* is defined as any deviation from an organism's normal, or "healthy," state (Ben-Haim and Rosenberg 2002; MedlinePlus 2003; Peters 2006; Stedman 2006). This includes impairment of vital functions, organs, or systems, including interruptions, cessation, proliferation, or other malfunctions, originating from either abiotic or biotic sources, or even combinations of two or more. These impairments are typically manifested through distinguishable signs (externally visible) and/or symptoms (felt internally by humans). "Infectious diseases are characterized both by an identifiable group of signs and the presence of the recognized etiologic or causative agent" (Ben-Haim and Rosenberg 2002). A summary of other important disease-related terminology is provided by **Table 2.1**.

Table 2.1 Important Disease Related Terminology

Contagious	Communicable or transmissible by contact with the sick or their fresh secretions or excretions
Disease	An interruption, cessation, or disorder of a body, system, or organ structure or function. A morbid entity ordinarily characterized by two or more of the following criteria: recognized etiologic agent(s), identifiable group of signs and symptoms, or consistent anatomic alterations
Endemic	Denoting a temporal pattern of disease occurrence in a population in which the disease occurs with predictable regularity with only relatively minor fluctuations in its frequency over time
Epidemic	The occurrence in a community or region of cases of an illness, specific health-related behavior, or other health-related events clearly in excess of normal expectancy
Epidemiology	The study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control health problems
Epizootic	An outbreak (epidemic) of disease in an animal population
Epizootiology	Epidemiology of disease in animal populations
Etiology	The science and study of the causes of disease and their mode of operation
Exposure	Proximity or contact with a source of a disease agent in such a manner that effective transmission of the agent or harmful effects of the agent may occur
Health	The state of the organism when it functions optimally without evidence of disease or abnormality
Incidence	The number of specified new events during a specified period in a specified population
Infectious	A disease capable of being transmitted from person to person, with or without actual contact
Morbid	Diseased or pathologic. The ratio of sick to well people in a community
Morbidity	Diseased state.
Mortality	A fatal outcome, synonymous with "death rate"
Pandemic	Denoting a disease affecting or attacking the population of an extensive region, country, continent, global
Panzootic	An epizootic occurring on a global scale
Pathogen	Any virus, microorganism, or other substance causing disease.
Opportunistic Pathogen	An organism that is capable of causing disease only when the host's resistance is lowered
Pathognomy	Diagnosis by means of a study of the typical symptoms of a disease, or of the subjective sensations of a patient
Pathology	The form of medical science and specialty practice concerned with all aspects of disease
Prevalence	The number of cases of a disease existing in a given population at a specific period of time
Sign	Any abnormality indicative of disease, discoverable on examination of the patient; an objective indication of disease
Symptom	Any morbid phenomenon or departure from the normal in structure, function, or sensation, experienced by the patient & considered to be a subjective indication of disease
Syndrome	Aggregate of symptoms and signs associated with any morbid process, together constituting the picture of disease

Note: the above definitions are taken directly from the 28th Edition of Stedman's Medical Dictionary (Stedman 2006)

2.2.2 Introduction to Disease Study

At the close of the 19th century, four scientists — Ignaz Semmelweiss, Louis Pasteur, Joseph Lister, and Robert Koch — created a benchmark in the study of disease. The Hungarian doctor Ignaz Semmelweiss (1818-1865) made the first breakthrough in disease etiology with his discovery of child-bed fever (pherperal fever), identifying the cause as a bacterial infection which doctors themselves were transmitting to their patients. He found that they could easily prevent this by simply washing their hands prior to subsequent patient contact (Burnet and White 1972; Bhopal 2002). While this may seem like common sense today, the concept of disease transport as the result of poor sanitation was a novel one at the time. Unfortunately for Semmelweiss, this idea was too radical for his peers who severely ridiculed him and drove him to a mental breakdown and consequent early death (Burnet and White 1972; Bhopal 2002).

The "germ theory of disease" was posed several years later by Louis Pasteur, "the father of modern bacteriology" (Conn 1895), who proposed that diseases were caused by micro-organisms (Conn 1895; McGill 2000). Pasteur developed his theory while working at a French vineyard where he found that wine souring was being caused by a foreign microbe, which could be killed by simply heating (pasteurizing) the wine (Conn 1895; McGill 2000). A man named Joseph Lister (for whom the popular mouthwash, Listerine, was later named) joined Pasteur, and, together, they discovered the "true nature of disease" (Conn 1895; Clark 1920; Ford 1928; Burnet and White 1972). In practice, they found that post-surgery infections could be greatly reduced by spraying phenol (the active ingredient in Listerine) on an open wound to kill any microbes present, and then bandage the wound sealing in the phenol and preventing any new microbes from coming into contact with the exposed internal tissue (Clark 1920; Ford 1928; Burnet and White 1972; McGill 2000).

Through his work on the bacterium that caused anthrax in 1876, Robert Koch became the first scientist to prove the microbial disease theory as posed by the three men before him (Burnet and White 1972; Bhopal 2002). Koch went on to write "Koch's Postulate" (**Table 2.2**) which set forth four rules that must be followed in order to definitively prove disease causation by a particular microbe (Burnet and White 1972). Before his death in 1910, Koch isolated the causative bacterial agents of tuberculosis (1882), conjunctivitis (1883), and cholera (1884) (Burnet and White 1972; Bhopal 2002; Jones 2004).

Table 2.2 Koch's Postulate (as defined by Burnet and White 1972; Balter 1998; Bhopal 2002)

In order to definitively state the cause of a disease as a specific microbe, the following rules must be adhered to:
1. The microbe must be present in all known cases of the disease, but not present in healthy (non-diseased) organisms
2. The microbe must be able to be isolated from the diseased organism and grown in pure culture in the lab
3. Experimental Infection: This lab grown microbe must cause the same disease when instilled in a healthy organism
4. The microbe must then be able to be isolated from the disease organism and grown in pure culture from the experimental infection in the lab

Koch's Postulate was fully accepted (and still is for many epidemiologists) until only recently when a sufficient number of studies were performed indicating that many causative agents could be identified with relative certainty but, because of their nature, would never adhere to the rules of the Postulate (Balter 1998; Van Gelder 2002). Raj Bhopal (2002), Professor of Public Health and Chair of the Department of Community Health Sciences at the University of Edinburgh, Scotland, refers to Koch's Postulate as "a counsel of perfection and too stringent even within the field of microbiology." Koch himself found exceptions to rules 1 and 3, in the form of asymptomatic carriers (Bhopal 2002). Despite this, over the past 25 years, marine microbiologists have found themselves confronted with a plethora of newly arising coral diseases, only a few of which would satisfy Koch's Postulate (Milius 1998).

2.2.3 Emerging Infectious Diseases (EIDs)

The Earth is currently experiencing an increase in Emerging Infectious Diseases (EIDs). These EIDs are targeting plants, wildlife (both terrestrial and marine), domestic animals, and humans resulting in an overall threat to biodiversity (Real 1996; Harvell et al. 1999,2002; Daszak et al. 2000; Harvell 2004). While Marine EIDs are rapidly increasing in both incidence and severity (Harvell et al. 1999,2002; McCallum et al. 2003; Harvell 2004), humans are not directly impacted by them as we are by the interactions among terrestrial EIDs. **Figure 2.1** illustrates the anthropogenic underpinnings of the causative factors and driving agents associated with these EIDs (Daszak et al. 2000; Harvell et al. 2002; Harvell 2004).

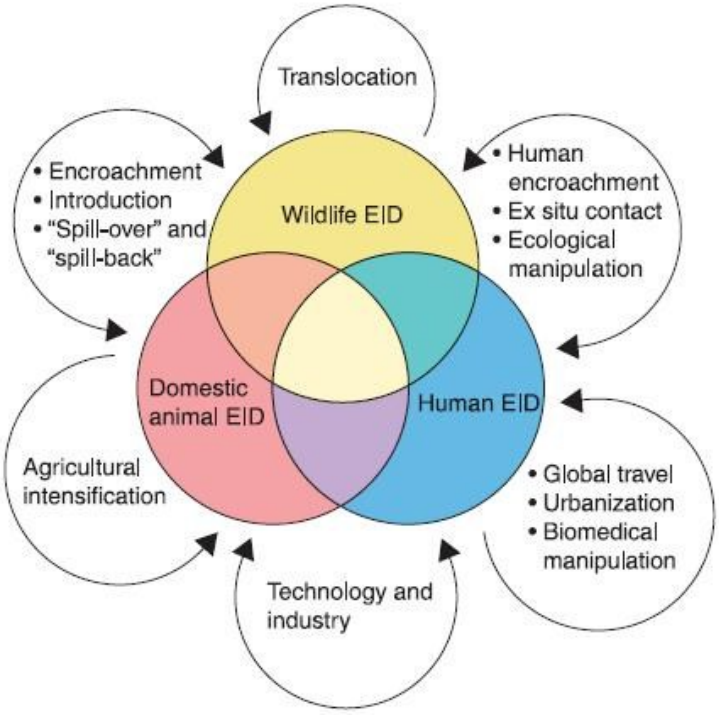


Figure 2.1 "The host-parasite ecological continuum (here parasites include viruses and parasitic prokaryotes). Most emerging diseases exist within a host and parasite continuum between wildlife, domestic animal, and human populations. Few diseases affect exclusively any one group, and the complex relations between host populations set the scene for disease emergence, Examples of EIDs that overlap these categories are canine distemper (domestic animals to wildlife), Lyme disease (wildlife to humans), cat scratch fever (domestic animals to humans) and rabies (all three categories). Arrows denote some of the key factors driving disease emergence." Note: both the diagram and legend were taken directly from Figure 1 on page 442 of Daszak et al. 2000.

Historically, we only have records of human EIDs, and more rarely domestic animal and agricultural EIDs which have a significant, usually economical, adverse effect to humans (Daszak et al. 2000; Harvell 2004). Environmental movements, and advancements in ecologic theory, have led scientists to expand their studies to non-economically profitable environmental systems; since environmental health is a function of all of Earth's ecosystems, not merely those profitable to humans. Daszak et al.'s (2000) theory of recent increases in EIDs are the result of increased human population levels. I will add to this that EID incidence is likely further increased by overall human ecosystem "modifications," more aptly described as intentional or unintentional anthropogenic environmental degradation (Western 2001). **Table 2.3** gives an accounting of the more noticeable human induced environmental degradations.

Since our transition from nomadic hunter-gathering tribes to agriculturally reliant civilizations, humans have modified the environment to better suit their "desires" or "whims." That is, humans have a tendency to act now and deal with the consequences later. The dramatic increase in EID incidence over the past few decades is therefore likely the result of centuries of our impulsive and irresponsible actions. Increased human population levels have caused further encroachment on the already substantially depleted remains of wildlife habitats; resulting in overall decreased wildlife populations, increased densities in remaining refuges and increased competition for space and resources, and overall quality of life, making them ripe for infection (Daszak et al. 2000; Harvell 2004). Removal of the barriers between wildlife, domestic animals, and humans, as well as the overall ramifications of human caused species globalization, has caused a recent, and seemingly unstoppable, increase in the "spill-over" and "spill-back" transmission of diseases between these three categories (Daszak et al. 2000; Harvell 2004; Power and Mitchell 2004)

Table 2.3 Anthropogenically Induced Environmental Changes

Characteristics of intentionally modified ecosystems

- High natural resource extraction
- Short food chains
- Food web simplification
- Habitat homogeneity
- Landscape homogeneity
- Heavy use of herbicides, pesticides, & insecticides
- Large importation of non-solar energy
- Large importation of nutrient supplements
- Convergent soil characteristics
- Modified hydrological cycles
- Reduced biotic & physical disturbance regimes
- Global mobility of people, goods, & services

Ecosystem side-effects of human activity

- Habitat & species loss (including conservation areas)
- Truncated ecological gradients
- Reduced ecotones (transition zone between ecosystems)
- Low alpha diversity
- Loss of soil fauna
- Simplified predator–prey, herbivore–carnivore, & host–parasite networks
- Low internal regulation of ecosystems due to loss of keystone agents
- Side effects of fertilizers, pesticides, insecticides, & herbicides
- Invasive nonindigenous species, especially weeds & pests
- Proliferation of resistant strains of organism
- New & virile infectious diseases
- Genetic loss of wild & domestic species
- Overharvesting of renewable natural resources
- High soil surface exposure & elevated albedo
- Accelerated erosion
- Nutrient leaching & eutrophication
- Pollution from domestic & commercial wastes
- Ecological impact of toxins & carcinogenic emissions
- Atmospheric & water pollution
- Global changes in lithosphere, hydrosphere, atmosphere, & climate

Some ecological consequences of human activity on ecosystem processes

- Ecosystem structure
- Loss of biodiversity
- Structural asymmetry & downsizing of communities
- Loss of keystone species and functional groups
- Ecosystem processes
- Low internal regulation
- High nutrient turnover
- High resilience
- Low resistance
- Low variability
- Low adaptability
- Ecosystem functions
- High porosity of nutrients & sediments
- Loss of productivity
- Loss of reflectance
- Global processes
- Modified biogeochemical cycles
- Atmospheric change
- Accelerated climatic change

Note: the above table was adapted from Tables 1, 2, and 3 on pages 5459 and 5461 of Western 2001

2.3 Diseases and the Marine Environment

2.3.1 Challenges Associated with Studying the Marine Environment

Perhaps one of the most important aspects in understanding the instigation of diseases in general, lies in understanding the crucial relationship between microorganisms and their hosts, since any deviation from their unique equilibrium relationship can result in the host taking on a diseased state (Dubos 1961; Rosenberg and Ben-Haim 2002). While terrestrial based diseases have been studied for centuries, marine diseases have remained for the most part an enigma to scientists (McCallum et al. 2003,2004). This is for a number of reasons, the first of which is that until the end of the 20th century the marine environment, outside of the intertidal zone, was for the most part off limits to researchers since it was largely inaccessible until the invention of SCUBA (Miloslavich et al. 2010; Dubinsky and Stambler 2011). The second, and perhaps more important, is that oceanic systems are much more complex than terrestrial systems, which makes pathogen transmission extremely difficult to detect, track, and study (McCallum et al. 2003,2004; Harvell et al. 2004). Because the area of marine studies is so recent, there is little baseline data in the field, which makes the recognition of altered states of health rather difficult as in many cases recorded states of normalcy do not exist (Harvell et al. 1999,2002,2004).

2.3.2 Historical Overview of Marine Diseases

Since the industrial revolution, there has been substantial decline in ecosystem health around the world (Revelle et al. 1965; Hardin 1968). In general, this decline has been caused directly or indirectly by anthropogenic stressors; such as: increased pollution levels, habitat manipulation and destruction; global warming; and general human over-population (Revelle and Suess 1957; Revelle et al. 1965; Hardin 1968; SCEP 1970; Hardin 1976; Kellogg 1987). The ramifications of our increasingly mechanized world became rapidly apparent and severe, resulting in the initial inklings of environmentalism (Carson 1962; Leopold 1966; Hardin 1968).

In the 1960s and early 1970s the environmental movement exploded onto the scene with the eye opening accounting of how humans had caused the widespread use of DDT such that virtually no part of the world had been left untouched (Carson 1962). This concept was further supported by Aldo Leopold's accountings of the changes and overall deterioration of nature that he witnessed during his lifetime, saying more than once that when one has had a beautiful interaction with nature it is best never to return to that spot, because it will undoubtedly be depleted and otherwise humanized (Leopold 1966).

It was during this same time period that SCUBA became more widely available to the public, allowing scientists to stay down longer and explore deeper realms of the ocean (Bascom and Revelle 1953; Dill and Shumway 1954; Kidson et al. 1962; Wood 1963; Fager et al. 1966; Roberts 1968; Richardson 1999). As this was a relatively new field, much of the initial observations in the ocean were descriptive in nature (Carson 1941,1951,1955; Doukan 1957; Thomson 1957; Link 1959; Goggin 1960; Schmid 1965; Macintyre 1967). It was during one of these descriptive dives that marine scientist Antonius (1973) reported the first incidence of coral disease. Over the past 30 years since that initial sighting, a plethora of coral diseases have been detected and studied (Gardner et al. 2003; Nowak 2004; Wapnick et al. 2004).

2.3.3 The Importance of Marine Health and Estimated Declines

Outbreaks of disease are known to modify the existing structure, function, and stability of marine ecosystems (Raymundo 2005). Of these, infectious disease outbreaks have been shown to be the most menacing, as they have the ability to dramatically decrease biodiversity through rapid population declines, local extirpations, and eventually species extinctions (Harvell et al. 2002). The severity of diseases is further confounded by the physical and chemical nature of the marine environment which not only enables, but may promote, the spread of marine pathogens (McCallum et al. 2003). Recent studies suggest that the macro-scale currents and lack of barriers

in the ocean enable long-distance dispersal of pathogens (McCallum et al. 2003). Currents, eddies, and strong winds most likely influence the directional migration of pathogens, in addition to further enhancing their rates of spread (McCallum et al. 2003). In fact, marine systems have been found to be capable of propagule dispersal rates more than two orders of magnitude greater than found on land (Kinlan and Gaines 2003; McCallum et al. 2003).

Despite the inherent problems and weaknesses with the study of marine diseases, they are especially important to study since the ocean accounts for 70.8% (Trujillo and Thurman 2008) of the Earth's surface , and tends to be a good indicator of environmental change—both natural and anthropogenically induced. These include increased sea levels, air and sea temperatures, nutrient levels, sedimentation, pollutants, and many others. Bruno et al. (2003) report that over the past 20 years there has been such a noticeable increase in both the prevalence and severity of marine diseases that a variety of keystone species are being significantly and adversely affected. In addition, commercially valuable, endangered, habitat forming foundation species are undergoing extreme reductions in both diversity and abundance as a direct result of these marine disease epizootics (Harvell et al. 1999; Bruno et al. 2003). Eventually putting the entire ecosystem at risk of wide scale change, such as that of their community structure, resulting most likely in instabilities in crucial ecosystem processes (Harvell et al. 2002; Bruno et al. 2003). Harvell et al. (2002) go on to describe infectious diseases as "strong biotic forces" due to their ability to cause cataclysmic population declines and species extirpations and extinctions. Of the many types of marine diseases it is especially important to study coral diseases as corals provide crucial habitat and geologic structure to coastal environments. Diseases tend to play a huge role in determining the structure of coral reef communities due to their rapid destructive abilities (Bythell and Sheppard 1993; Richmond 1993; Aronson and Precht 1997; Cooney et al. 2002).

Disease incidences in the marine realm have increased over the last few decades (Harvell et al. 1999; Ginsburg 2000; Aronson and Precht 2001; Daszak et al. 2001; Harvell et al. 2001; Precht et al. 2002), with a marked acceleration of reported instances from the 1900's to the present (Richardson 1998; Green and Bruckner 2000). While scientists lack baseline data and an overall epidemiological history for the ocean, recent studies are showing that the increased reports of disease occurrence is novel, and not purely an artifact of increased awareness and study of the marine environment (Harvell et al. 1999,2002,2004). This consensus was made after a period of global ecological monitoring of the oceans, in conjunction with paleontological studies which confirmed that while there were diseases in the past which affected the marine environment, the incidence, severity, and rates of spread, have risen dramatically in recent years and are not comparable to any historical outbreaks (Harvell et al. 1999; Aronson and Precht 2001; Porter et al. 2001; Bruno et al. 2003).

What is perhaps the most alarming about this increase is the degenerative, and in some cases eradicating, effects disease has had on various commercially valuable, foundation, habitat forming, keystone, as well as already threatened or endangered marine species (Harvell et al. 1999; Green and Bruckner 2000; Bruno et al. 2003). As of 2000, as many as thirty-four mass mortalities had been reported in a wide variety of marine groups, each affecting more than 10% of the infected population (Harvell et al. 1999; Green and Bruckner 2000). Perhaps the most severe, or at least most studied and publicized, of these outbreaks has been within corals.

Many scientists believe that corals may serve as an indicator group for the global marine ecosystem decline (Green and Bruckner 2000; Barber et al. 2001). The decline, and in some areas, collapse, of coral reefs around the world is best marked by five crucial biological responses within the reef system: (1) the absence of large marine animals, such as turtles, sharks, and groupers which had once been commonplace there (Jackson et al. 2001; Pandolfi et al. 2003,

2005); (2) a general decrease in the diversity of fish and invertebrates present on the reefs (Jackson et al. 2001; Pandolfi et al. 2003,2005); (3) mass bleaching events (Goreau et al. 1998; Barber et al. 2001; Harvell et al. 2001); (4) macroalgal overgrowth and overall community phase shifts (Goreau et al. 1998; Aronson and Precht 2001; Barber et al. 2001; Precht et al. 2002); and (5) emerging coral diseases (Goreau et al. 1998; Barber et al. 2001).

2.3.4 Potential Causes of Declining Marine Health

Right now there is a debate among the scientific community as to the cause of the increased marine disease incidence. Are the pathogens becoming more virulent than in the past? Are the immune systems of marine organisms weakening and becoming more vulnerable to infection? Are pathogens expanding out from their previously known ranges, and if so is this range expansion human-facilitated? Or lastly, have human activities stressed marine organisms so much that their susceptibility to infection has increased or prompted the outbreaks of these diseases (Harvell et al. 1999; Green and Bruckner 2000; Hayes et al. 2001; Harvell et al. 2002; Bruno et al. 2003)? The most likely answer is a combination of all of the above factors and stressors, resulting in the overall increased susceptibility of marine organisms to infection.

The timing of these outbreaks correlates with the present exponential human population growth causing many scientists to more closely examine the anthropogenic role in these outbreaks. Global increase in human populations along the coasts, has no doubt, caused the current increase of sewage discharge into the oceans (O'Shea and Field 1992; Harvell et al. 1999). In addition, sewage is making its way into the oceans via a number of different methods, both direct (e.g. storm water overflow, and direct dumping from cruise ships, fishing vessels, commercial ships, etc.) and indirect (underground seepage from septic tanks and injection wells). For this reason, it is not surprising that scientists are calling for increased management of water quality, specifically designed to limit pollution (Bruno et al. 2003).

With the oceans covering so much of the world's surface, why is it that disease incidence has not been more diffused over this vast area? The answer likely lies in the physical properties of the marine environment which appears to both promote pathogen survival (Colwell and Huq 2001; McCallum et al. 2003), and facilitate the potential for long-distance dispersal, which results in increased rates of pathogenic spread (McCallum et al. 2003). There is also a fundamental difference in how terrestrial and marine systems handle the physics behind dispersal of organisms and the degree of connectivity within the given ecosystem. This suggests that the underlying physics of the marine realm may facilitate much faster rates of pathogenic spread than possible on land (Kinlan and Gaines 2003; McCallum et al. 2003). Kinlan and Gaines (2003) used both direct and genetic methods to compare the differences in propagule dispersion for both land and water. Their results indicated that marine propagule dispersal occurred at rates as high as two orders of magnitude greater than the fastest terrestrial mode of dispersion, which they identified to be the dispersal of terrestrial plants. McCallum et al. (2003) explains this by the lack of dispersal barriers present in the marine realm thus facilitating more rapid pathogenic transport than on land.

There are undoubtedly a plethora of factors (natural or anthropogenic in origin) contributing to the current rise in marine EIDs, which are reacting in complex and possibly even synergistic ways that science is just starting to discover (Harvell et al. 1999,2004; Green and Bruckner 2000). To facilitate the study there are several research avenues which need to be pursued, among which understanding the role of climate change is perhaps most crucial, to the marine environment (Bruno et al. 2003). While scientists are just beginning to understand both climate change and its marine ramifications, there is evidence that temperature increase, sea level rise, changes in ocean circulation, and decreased salinity will directly impact the health of the marine ecosystems (Harvell et al. 2002; Bruno et al. 2003).

Chapter 3. Coral Reefs and Their Declining Health

"Corals form the structural and biological framework of some of the most diverse, productive and economically important marine ecosystems in the world. There is growing evidence that these ecosystems are now being degraded at an alarming rate as a result of the synergistic impacts of over-fishing, anthropogenically derived increases in carbon dioxide levels, warming of sea surface temperatures, eutrophication, sedimentation, and pollution (Lesser 2004)." – Lesser et al. (2007)

3.1 Introduction to Coral Reefs

In addition to being one of the oldest ecosystems in the world, dating back hundreds of millions of years (Wood 1998; Hoegh-Guldberg 1999; Pandolfi 2011), coral reefs are also one of the most complex, productive, and biologically diverse ecosystems in the world (Hoegh-Guldberg 1999; Lesser 2011; Weil and Rogers 2011). Additionally, coral reefs have provided "millions of people with food, building materials, protection from storms, recreation, and social stability over thousands of years, and more recently, income, active pharmacological compounds, and other benefits" (Weil and Rogers 2011). However, anthropogenic changes to earth's natural climatic, terrestrial, and oceanic systems have caused such a dramatic decline in reef health, that these "iconic" ecosystems are now not only considered threatened worldwide, but recent studies are showing that tropical coral reefs may disappear within the next 50 years (Anthony et al. 2008; Silverman et al. 2009; Veron et al. 2009; Erez et al. 2011; Hoegh-Guldberg 2011).

3.2 Coral Reef Ecology and Biology

The term "coral reef" is unique in that it refers to both a geological structure and a biological community (Buddemeier et al. 2004). The geological structure of the reef is essentially the limestone (calcium carbonate) framework, which is made by the biologic organisms (namely corals and calcareous algae) which secrete the calcareous material as part of their growth. The biological use of the term "coral" refers to the coral animals (Phylum

Cnidaria) that make up the living part of the reef (Buddemeier et al. 2004; Harrison 2011). The term "reef," in a biological context, generally refers to the whole ecosystem, from the corals to the fish and countless other organisms, which are associated with reefs. The above terminology is important, as there is often much confusion as to what is being referred to when the use of the terms are not qualified appropriately.

3.2.1 Reef Formation and Zonation

Limestone is a sedimentary rock composed of the minerals calcite and aragonite, which are different forms of calcium carbonate ($CaCO_3$; Ohde and Hossain 2004). Corals create their calcareous skeletons by secreting calcium (Ca^{2+}) and bicarbonate (HCO_3^-) ions through the outer layer of tissue on the underside of the polyps (Gattuso et al. 1999; Ohde and Hossain 2004). These secreted ions form a matrix of $CaCO_3$, providing an additional layer to the underlying skeletal structure (Ohde and Hossain 2004). The calcification process is very energy demanding because the ions must be propelled out against a gradient (Gattuso et al. 1999). The majority of the energy needed for this process (up to 95%) is provided by the Zooxanthellae that live within the coral tissue (Gattuso et al. 1999; Hoegh-Guldberg 1999). Zooxanthellae are unicellular, photosynthetic dinoflagellates (algae) that have a symbiotic relationship with their coral host (Hoegh-Guldberg 1999). Because this process is fueled by the energy provided by the photosynthetic algae, it is often referred to as "light-enhanced calcification" (Vandermeulen et al. 1972; Chalker and Taylor 1975). The pH levels are an important component to the skeleton formation, because the skeleton dissolves in acidic solutions (low pH). For this reason, corals must maintain a high pH concentration (Gattuso et al. 1999). These levels are the highest during the day ($pH \approx 9.3$), dropping to a pH of roughly 8.0 at night (Kleypas et al. 1999a; Kleypas et al. 1999b; Ohde and Hossain 2004).

The geomorphology and subsequent zonation of shallow-water reefs is primarily shaped by the combination of water-temperature, light penetration, and prevailing wave energy (Bak 1983; Charuchinda and Hylleberg 1984; Hubbard 1997). These factors do not usually have a uniform distribution, but rather change with depth, distance from shore, and type of shore habitat (sandy, rocky, mangroves, etc). Over time, these differences cause different zones within the reef to develop (**Figure 3.1**). Eventually organisms will begin to dominate the reef zone in which they are best suited, thus creating biological reef zonation (**Figure 3.1**).

When hurricanes are added into the mix, different combinations of the intensity and frequency of both waves and storms results in the formation of the following three overall types of Caribbean reefs: Pavement reefs, which have low ambient wave conditions and high hurricane frequencies; Algal Ridge reefs, which have high prevailing wave conditions and high hurricane frequencies; and *Acropora palmata* dominated reefs (**Figure 3.1A**), which have high wave energies and experience hurricanes less frequently than the previous two reef types (Hubbard 1997; Hubbard et al. 2008). Pavement and Algal Ridge reefs have few, if any, corals, due in large part to the high hurricane frequency. Hurricanes and strong storms are known to cause the branches of *A. palmata* to break off. High wave energy prevents nutrients, sediments, pollution, and potentially other stressors from accumulating on and around the reef, thus lowering the stress levels of the coral. If the coral is relatively healthy, then the high wave energy will also help to encourage coral recruitment and the reef will rebuild and recover.

3.2.2 Caribbean Reef Distribution

Hermatypic ("stony" or "reef-building" corals), coral reefs are generally characterized as "steno-tolerant" ecosystems, because there is a relatively narrow ("steno") range of environmental conditions under which they are known survive, much less thrive (Kleypas et al. 1999a). Specifically, hermatypic corals are usually limited to "warm, clear, shallow, and fully

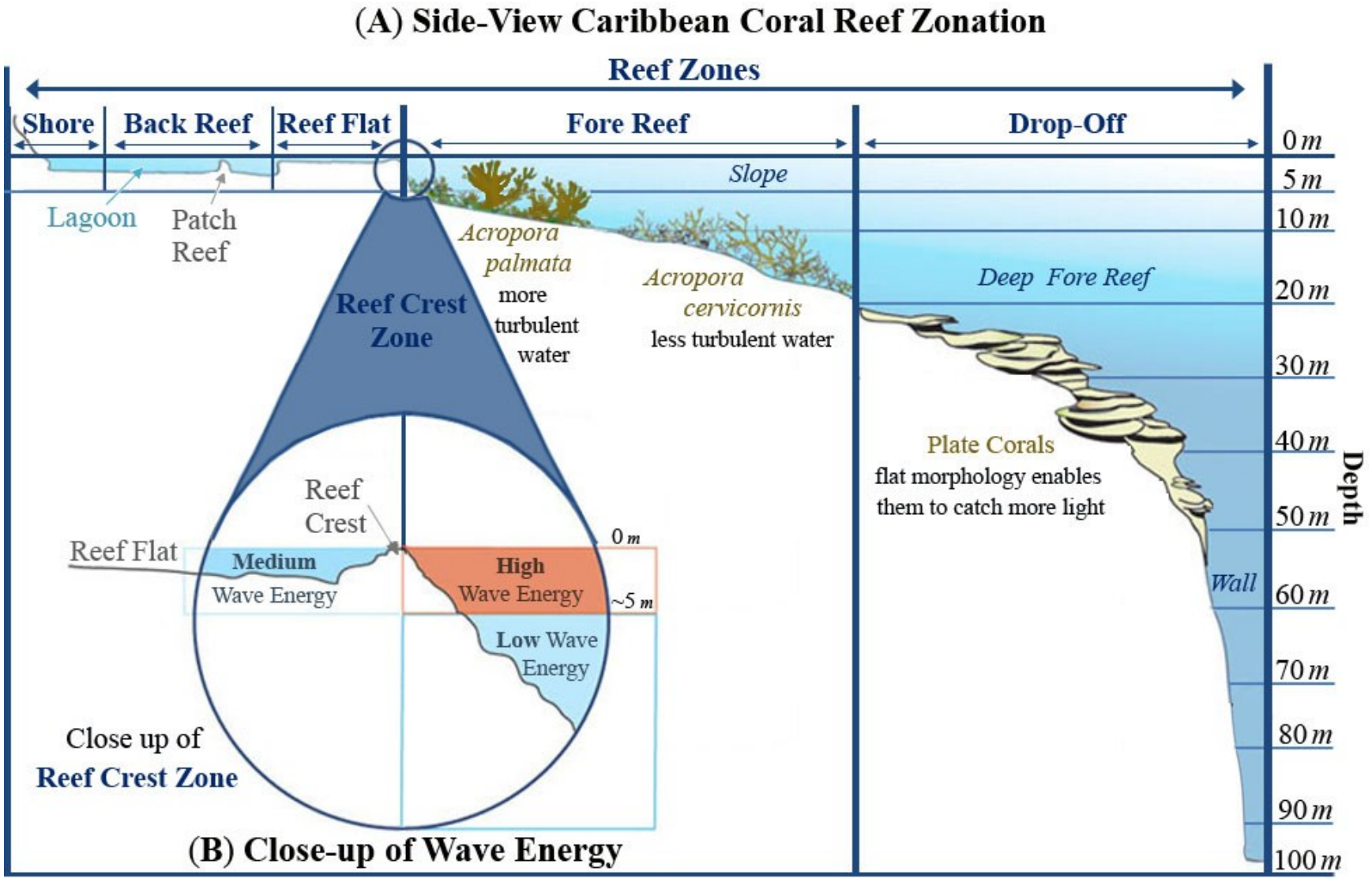


Figure 3.1 Reef Zonation. (**A**) A side-view of some of the common types of reef zones found in the Caribbean. (**B**) A close-up of the different types of wave energy generally found on either side of the reef crest. Note: this figure was adapted from the diagram on the following website: *http://media.beautifuloceans.com/course1/pic/1.1_CoralReef_Zonation_800pix.jpg*.

saline waters" (Kleypas et al. 1999a). Given this limited range of environmental conditions, the geographic distribution of most coral reefs is usually confined to tropical and subtropical, shallow waters located within 30 latitudinal degrees north or south of the equator (Birkeland 1988; Dubinsky and Falkowski 2011; Hoegh-Guldberg 2011; Lesser 2011; Pandolfi 2011). The equatorial region has high angle of solar irradiance year round, resulting in warm waters with little temporal fluctuation throughout the year (Miloslavich et al. 2010). For example, the surface waters in the Caribbean generally range from 22°C – 29°C (Miloslavich et al. 2010), with the exception of the Gulf of Mexico basin which is considerably cooler in the winter months.

Corals are generally found in this zone because hermatypic corals have a relatively narrow thermal range and historically temperature has proven to be one of the primary environmental factors controlling not only their geographic distribution, but their very survival as well (Dana 1843; Mayor 1914,1915; Vaughan 1918,1919; Vaughan and Wells 1943; Walker et al. 1982; Glynn and D'Croz 1990; Kleypas et al. 1999a; Lirman et al. 2011). While thermal tolerance is known to vary by species and location (Walker et al. 1982; Berkelmans and Willis 1999; Marshall and Baird 2000), the minimal and maximal critical thermal stress thresholds of most corals are defined as 16°C (60.8°F) and 36°C (96.8°F), by Mayor (1915) and Kinsman (1964), respectively. However, most hermatypic corals live within a thermal range of 25°C – 29°C (Vaughan and Wells 1943; Kinsman 1964; Jaap 1979; Walker et al. 1982), with an optimal growth occurring between 26ºC – 27°C (Clausen and Roth 1975; Jokiel and Coles 1977; Marshall and Clode 2004; Lirman et al. 2011).

In addition to temperature, there are other physical and environmental properties that influence the geographic distribution of corals; such as: salinity, light, nutrients, sediment, depth, and various hydrodynamic factors (Sheppard et al. 2009). For example, coral reefs are more

likely to occur on the eastern side, rather than western side of continents within their equatorial zone. Large scale oceanic gyres, driven by wind systems and the earth's rotation, produce conditions unfavorable to reef production on western sides of continents or eastern sides of ocean basins (Walton Smith 1971). This results from local wind-driven upwelling of cold nutrient-rich water from below as well as equatorial flow of colder water from high latitudes (i.e. eastern boundary currents; Walton Smith 1971). Regions with upwellings are not conducive to reef formation, as the waters are too cold, nutrient rich, and have poor water clarity; consequently any reefs that do form in these areas are generally "less extensive and more fragmentary" (Sheppard et al. 2009). Conversely, the warm, western boundary currents found on the eastern side of continents generally provides a much more favorable environment for corals.

The presence of warm, surface currents also play an important role in the reproductive success and geographic distribution of coral reefs by transporting coral planulae (larvae) throughout the region. The velocity of the currents may also help reduce the impact of different stressors. For example, fast currents can lessen the impact of bleaching by moving the warm water along and preventing prolonged exposure (Grimsditch and Salm 2005). Fast currents also help reduce algal stress by inhibiting the settlement of macroalgae, thus allowing coral recruits to settle and grow (Grimsditch and Salm 2005). Last, currents generated by breaking waves are also important to reefs because they transport essential nutrients throughout the system, as well as, moving waste products out of the area (Hubbard 1997).

3.2.3 Phylogenic and Taxonomic Classification of Corals

According to Woese et al.'s (1990) three-domain phylogenic classification system, corals are classified as eukaryotes (Domain "Eucarya") because they have cells with a nucleus containing their genetic material (Sumich and Morrissey 2004). Within this domain, corals are classified as animals (Kingdom Animalia) according to Whittaker's kingdom-based taxonomic

classification system (Whittaker 1959,1969; Whittaker and Margulis 1978), because they are multicellular eukaryotic organisms whose cells lack cell walls and have some degree of muscle-contracting and nerve-conducting capabilities (Sumich and Morrissey 2004). Within this kingdom, corals are classified as animals that belong to the Phylum Cnidaria (see **Figure 3.2**). This phylum is an exclusively aquatic group of carnivorous organisms that have radially symmetrical, simple body structures, in which the mouth is the only opening and it is surrounded by nematocyst-lined tentacles that aid in capturing their prey (Sumich and Morrissey 2004). Cnidarians generally exist as either free-swimming medusa (jellyfish) or sessile polyps attached to the benthos (sea anemones and corals). Corals are differentiated from other cnidarians, because their life-cycle is dominated by the polypoid generation as opposed to jellyfish whose life-cycles are dominated by the medusoid generation (Sumich and Morrissey 2004).

Corals are classified as either Anthozoans or Hydrozoans (Class Anthozoa or Hydrozoa, respectively), based on their morphologic form (polypoid and/or medusoid) and the anatomic structure of their gastervascular cavity and gastrodermis (see **Figure 3.2**). Hermatypic (reef-building) corals are generally considered to be Scleractinian anthozoans (Order Scleractinia), characterized by their hard, calcareous exoskeleton and photosynthetic endosymbionts (zooxanthellae). Other corals that are known contribute to the reef-building process are usually not considered to be *hermatypic* because they either lack zooxanthellae or are not Scleractinians.

3.2.4 Basic Anatomy of Hermatypic Corals

Corals are an especially interesting marine animal as there is very little about them that is animal-like, but rather the organisms can be better understood by thinking of them as plants. Corals are colonial animals made up of thousands of connected polyps (**Figure 3.3B**, **C**), in which each polyp represents an individual coral animal (**Figure 3.3C**, **D**). Hermatypic coral polyps have a soft, sea anemone-like body that is protected by the corallite (**Figure 3.3D**). The

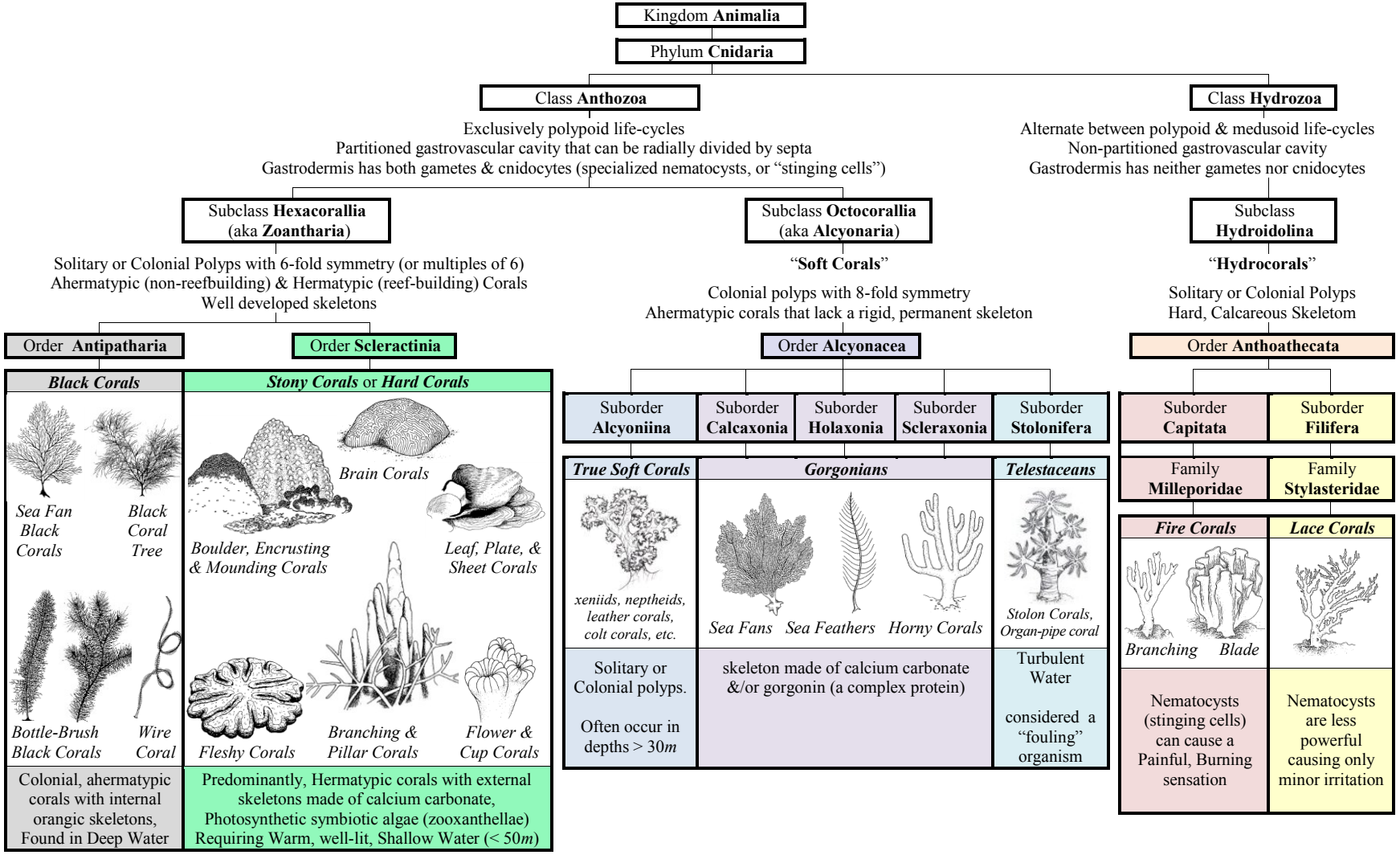


Figure 3.2 Taxonomic classifications, characteristics, and depictions of common types of corals. Note: the above taxonomic classifications and characteristics were based on the following sources: Lutz (1986); Fautin & Romano (1997,2000); Fautin et al. (2000); Humann and Deloach (2002); Romano and Cairns (2002); Rose (2009); Sheppard et al. (2009); OBIS (2011). The illustrations are from Humann & Deloach's (2002) "Reef Coral Identification: Florida Caribbean Bahamas" book.

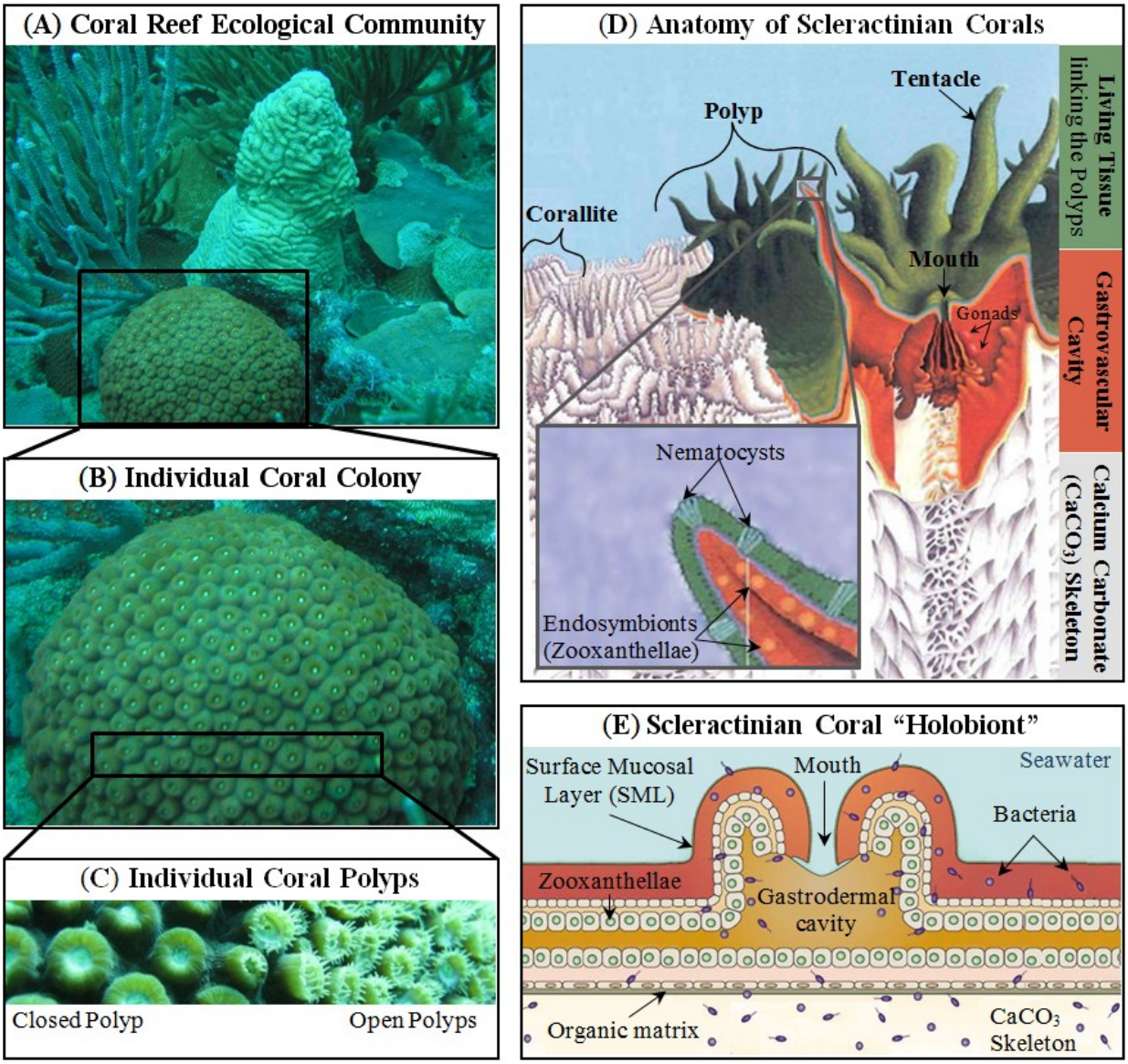


Figure 3.3 Coral reefs and the basic anatomy of hermatypic corals. (**A**) An example of a coral reef ecological community. (**B**) This is a close-up of the *Montastrea cavernosa* coral shown in **A,** representing an individual coral colony. (**C**) A close-up of *M. cavernosa* polyps depicting the difference between the visual appearance of an open polyp (right) and a closed polyp (left). (**D**) A cross-sectional diagram depicting the major anatomical elements associated with scleractinian, hermatypic corals. (**E**) Cross-sectional diagram of the coral holobiont, depicting the some common microbial inhabitants in relation to the anatomy of a scleractinian polyp. The microbes are not drawn to scale. Note: the photograph shown in **A** (and **B**) was taken in Puerto Rico by NOAA's Biogeography Team, Center for Coastal Monitoring (*http://www8.nos.noaa.gov/biogeo_public/habitat_photos.aspx*). The photograph of the *M. cavernosa* polyps shown in **C** was taken by Steve Vollmer's lab, and is available online at: *http://nuweb5.neu.edu/vollmerlabwp/category/potential-students/*. The diagram shown in **D** was adapted from Geoff Kelly's figure in Veron (2000), and the diagram shown in **E** was adapted from Figure 1, on page 356 of Rosenberg et al. (2007b).

term “corallite” refers to the calcareous skeleton of an individual scleractinian polyp (James 1974).

The coral grows upwards and outwards by secreting calcium-carbonate ($CaCO_3$) at its base, this occurs at the interface between the polyp (green) and the underlying skeleton (white) depicted in **Figure 3.3D**. The rate of coral growth varies according to the type and health of the coral. Branching corals, such as the two types of *Acropora* corals depicted in the Fore Reef Zone of **Figure 3.1A**, may grow as much as 15*cm*/*year*; while massive corals, such as the upper row of scleractinian corals depicted in **Figure 3.2**, may only grow 1-2*cm*/*year* (Spalding and Bunting 2004). On average (excluding the rapidly growing reef-building corals) it takes a thousand or more years for a coral reef (**Figure 3.3A**) to grow one *meter* (~3.28 *feet*); the individual coral colonies (**Figure 3.3B**) generally grow an order of magnitude faster than this adding the same height in roughly one hundred years (Spalding and Bunting 2004). However, when corals become stressed corals, they must focus their energy on surviving rather than growth or reproduction (Szmant and Gassman 1990; Lirman 2000; Fine et al. 2001; Rosenberg and Barash 2005).

3.2.5 The “Holobiont”

The “coral holobiont” is defined as the entire community of living organisms that make up a healthy coral head (Rohwer et al. 2002; Rosenberg et al. 2007a,b; Siboni et al. 2008; Bourne et al. 2009). The holobiont contains both macro- and micro-organisms from all three domains, as well as viruses which don’t really fall under any of the three domains (**Figure 3.4**). The primary Eukarotic members of the holobiont (**Figure 3.3E**) are the coral animal itself, and the microscopic endosymbiotic dino-flagellates (algae) commonly referred to as “zooxanthellae”. Other Eukarotic members of the coral holobiont include alveolates, basal protists, chromists, endolithic algae, flagellates, and fungi. The micro-organismal members from the Bacteria

domain include cyanobacteria and various types of endolithic, endosymbiotic, ectosymbiotic, and heterotrophic bacteria (Rosenberg et al. 2007a,b; Bourne et al. 2009). The members from the Archea domain are still being studied, but so far appear to predominantly come from the Crenarchaeota and/or Euryarchaeota phyla (Kellogg 2004; Wegley et al. 2004; Rosenberg et al. 2007a). In order to understand the role that the members from each of these domains plays in the susceptibility to diseases (including bleaching), the role they play within the complex structure of the holobiont (**Figures 3.3E** and **3.4**) must first be examined.

Corals are colonial animals made up of thousands of connected polyps (**Figure 3.3B**, **C**). Lining the interior of the polyp are a layer of microscopic algae, zooxanthellae (**Figure 3.3D inset** and **3.3E**), which give corals their *plant-like* nature (Buddemeier et al. 2004). The zooxanthellae and the coral polyps have a symbiotic relationship in which the algae use CO_2 as fuel to perform photosynthesis in order to produce food. The algae pass this food onto the corals in exchange for carbon dioxide, nutrients, and shelter. To give an idea of the size difference between the coral host and its endosymbionts, one coral polyp has ~1-5 million zooxanthellae per cm^2 (Spalding and Bunting 2004).

The bacterial members of the holobiont are usually found within the surface mucus layer (SML), the coral tissue, and the porous calcium carbonate ($CaCO_3$) skeleton formed by their Cnidarian host (**Figure 3.3E**; Harvell et al. 2007; Rosenberg et al. 2007a,b). In some cases bacteria appear to be associated with disease resistance by doing one (or a combination) of the following: producing antibiotics; taking up space; and/or forming antagonistic relationships with other bacteria in order to prevent one, potentially pathogenic, bacterial community from taking over the mucosal zone (Riegl et al. 2009; Rypien et al. 2010). However, there are numerous other studies suggesting bacteria are what is causing the disease (Kushmaro et al. 1996-1998, 2001; Smith et al. 1996; Nagelkerken et al. 1997; Geiser et al. 1998; Richardson et al. 1998a,b;

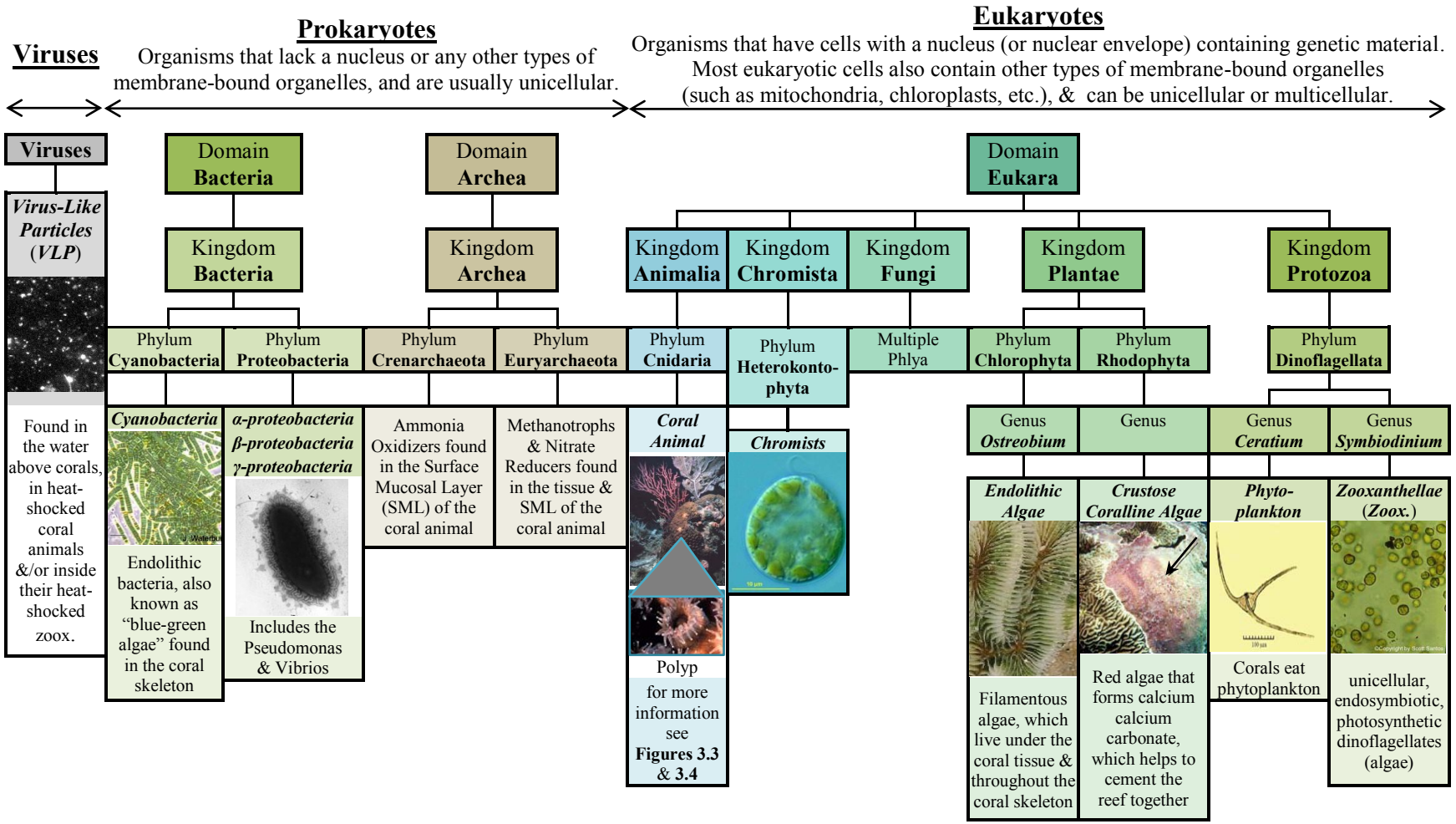


Figure 3.4 Some of the organisms (both macro- and micro-) known to make up the coral Holobiont (Shashar et al. 1997; Baker 2003; Kellogg 2004; Wegley et al. 2004; Wilson et al. 2005; Beman et al. 2007; Rosenberg et al. 2007a; Siboni et al. 2008; Lins-de-Barros et al. 2010; OBIS 2011). Note: the virus and endolithic algae images were taken from Figure 1 on page 147 of Rosenberg et al. (2007b); all other images are from NOAA's Coral Reef Information System (CoRIS) website (*www.coris.noaa.gov/*).

Ritchie and Smith 1998; Smith et al. 1998; Ben-Haim and Rosenberg 2002; Patterson et al. 2002; Ben-Haim et al. 2003a,b; Denner et al. 2003; Sutherland et al. 2004,2010-2011; Harvel et al. 2007). One study shows that the coral's skeleton may protect the holobiont by absorbing strong ultraviolet radiation, that is found to be detrimental to most other marine life (Reef et al. 2009).

The Archaea microbial members tend to be associated with the SML of the polyp, where they are thought to contribute to the nitrogen cycle by acting as a sink for the excess ammonium the ends up being trapped within the mucus layer during the nitrification and denitrification processes (Siboni et al. 2008). However, there is still so little know about the Archea that it is unclear whether or not they might play a role in disease susceptibility (Siboni et al. 2008).

Over the last decade, numerous microbial studies have greatly enhanced the scientific understanding of the coral holobiont (Rohwer et al. 2001,2002; Kellogg 2004; Wegley et al. 2004; Bourne and Munn 2005; Johnston and Rohwer 2007; Rosenberg et al. 2007b; Chimetto et al. 2008; Siboni et al. 2008; Ainsworth and Hoegh-Guldberg 2009; Reis et al. 2009; Lins-de-Barros et al. 2010). Much of this work has been focused on studying how the holobiont deals with stress (Reshef et al. 2006; Thurber et al. 2009). Progress in this area has been slow, especially given the complex nature of the relationships between and among the flora and fauna associated with the coral holobiont (Ainsworth and Hoegh-Guldberg 2009). This is further complicated by the lack of baseline data on healthy coral holobionts (Klaus et al. 2005; Lins-de-Barros et al. 2010). In fact, the very concept of the "holobiont" with regard to corals is still relatively new, prior to this past decade most microbial studies of corals were focused more on the relationship between the zooxanthellae and the coral host, than on identifying and/or understanding the roles of the other microbes present (Lins-de-Barros et al. 2010).

Increasingly studies are showing that corals worldwide are experiencing stress from a variety of sources. Consequently, studies of apparently healthy coral holobionts, may actually be

characterizing the inhabitants and functions of holobionts under low-grade stress (Casas et al. 2004; Bourne and Munn 2005; Lins-de-Barros et al. 2010); thus, making it difficult to differentiate between healthy and diseased corals based on the presence or absence of specific microbes and whether or not the presence of these microbes is beneficial to the well-being of the coral. Without a firm understanding of the microbiota of healthy corals it has proven difficult, and in many cases impossible, to confirm whether or not the bacteria (and other microbes) associated with coral lesions were responsible for causing the lesions (Bourne et al. 2009). Despite the numerous difficulties associated with determining and confirming disease etiology at the microbial level, most scientists agree with the following statement by Rene Dubos (1961):

> In general, the adaptive relationships between microorganisms and their hosts is effective only for the precise circumstances under which adaptation evolved... Any departure from this normal state is liable to upset the equilibrium and bring about a state of disease.

3.3 Coral Health and Disease

3.3.1 Significance of Reef Health

As indicated by their nickname "canaries of the sea" (Gustavson et al. 2000), coral reefs are an indicator of larger marine ecosystem health. Because of the small, sensitive, ecological niche in which they live (Kleypas et al. 1999a), minute changes in their habitat can trigger large scale community change within the reef. Thus, when the reefs are thriving the oceans seem to be in harmonious balance; however, when reef ecosystems begin to crash, as they are now, it is a sign of larger environmental problems, which left unchecked, will have a "domino" effect on the surrounding marine environment.

Despite only occupying one tenth of one percent of the world's oceans, 284,300km^2 (Spalding et al. 2001; Downs et al. 2005; Hoegh-Guldberg 2005), coral reefs contribute substantially to not only the marine environment, but the biological communities, physical and

atmospheric processes of the entire globe (Hoegh-Guldberg 1999). Ecologically, coral reefs are one of the most specious and diverse ecosystems in the world, supporting more than one million marine species (Hoegh-Guldberg 1999). Coral reefs provide essential habitat, shelter, and breeding grounds for the world's greatest biodiversity (Wood 1998; Donia and Hamann 2003; Bellwood and Meyer 2009).

Physically, coral reefs serve as the first line of defense against shoreline erosion (Wood 1998; Hoegh-Guldberg 1999). In the past the reefs protected the coasts from waves, floods, and tropical storms (Wood 1998; Hoegh-Guldberg 1999; Lirman 1999). Today, the coastal protection provided by the reefs is crucial as sea levels rise, severe weather increases in frequency, and overall climatic extremes intensify due to human induced climate change (global warming).

Chemically, oceans are a sink for carbon dioxide, helping to draw the greenhouse gases out of the atmosphere (Hoegh-Guldberg 1999). Reefs are also currently being explored by pharmacologists in search of miracle drugs, such as AZT derived from a Caribbean reef sponge and used to treat HIV patients, cancer treatments, antiviral medications, and painkillers (Bruckner 2002b; Mescher and Sturgess 2009).

Economically coral reefs support a large tourism industry. Caribbean reefs are a multi-billion dollar tourism industry (Hoegh-Guldberg 1999). One study estimated the total annual revenue from reef related tourism in Australia to be on the order of 4.3 billion Australian dollars; a stark contrast to the 360 million Australian dollars generated by commercial and recreational fishing (Butler 2005). Both the act of fishing and the depletion in fish stocks are stressful to reef communities, the Australian government recently made more than one-third of the Great Barrier Reef Marine Park "no take" zones (Evans and Russ 2004; Butler 2005); as a healthy reef proved to be FAR more profitable than a fledgling fishing industry. Buddemeier et al.'s (2004) study

estimated the annual net economic benefits of the world's coral reefs at more than 30 billion dollars.

Currently more than 10% of the world's population live within 100*km* of coral reefs (Donner and Potere 2007). The vast majority of these people live in undeveloped or developing countries, and rely heavily on the reefs for both their food and livelihood (Hoegh-Guldberg 1999; Andrefouet et al. 2005; Donner and Potere 2007). Seventy-five percent of these people live in the poorest developing countries (Donner and Potere 2007). For most island nations, fisheries and aquaculture are the primary food source, since the land is too scarce to rely on agriculture (Hoegh-Guldberg 1999; Donner and Potere 2007). In many parts of the South Pacific subsistence fishing accounts for 80% of the catch; unfortunately, subsistence fishing is not included in national fishing statistics (Donner and Potere 2007), resulting in overestimations of the fisheries and ultimately their collapse. If coral reefs collapse, not only will these people not be able to feed and support themselves, but the resultant shoreline erosion coupled with increasing water levels will leave many of these people homeless.

3.3.2 Historical Overview of Coral Diseases

As marine ecology is a relatively new discipline, baseline data on coral health, and disease distribution and abundance is sparse, especially when compared to terrestrial studies (Harvell et al. 1999). Many critics dismiss the notion of an anthropogenic link between increased disease incidence, abundance, and severity on corals, arguing that diseases have likely always around, we just were not looking for them before. While it is likely that corals have battled diseases in the past, evidence is increasingly showing that various stressors, both natural and anthropogenic in origin, are acting synergistically resulting in wide spread decimations of coral populations and in the severe cases species extirpation (Bruno et al. 2003; Sutherland et al. 2004; Wapnick et al. 2004). As the reports of new diseases, outbreaks, and population

decimations continue to rise, scientists are increasingly turning to paleontologic studies for insight on how unique these occurrences were on the geologic scale, so as to attempt to narrow down the realm of possible causes for the destruction (Porter and Meier 1992; Grigg 1994; Hughes 1994; Jackson 1997,2001; Cooney et al. 2002; Garrison et al. 2003).

Unfortunately, this question cannot easily be resolved, as there are disagreements within the paleontologic community on the issue. Pandolfi et al. (2003) reports that analysis of records extending back thousands of years shows that "degradation of coral reef ecosystems began centuries ago" and that "all reefs were substantially degraded long before outbreaks of coral disease and bleaching." While other scientists do not disagree with Pandolfi's assessment that the reefs were severely degraded hundreds of years before we began to study them. Instead, these scientists believe that recent coral declines are substantially more severe and wide spread, even if their presence is not novel (Hubbard et al. 1994; Bruckner and Bruckner 1998; Epstein et al. 1998; Harvell et al. 1999; Kim et al. 2000a,b; Aronson and Precht 2001; Porter et al. 2001; Rosenberg and Ben-Haim 2002; Bruno et al. 2003; Garrison et al. 2003; Sutherland et al. 2004; Wapnick et al. 2004); further arguing that this recent increase in both incidence and severity is not an artifact of increased attention (Epstein et al. 1998; Goreau et al. 1998; Harvell et al. 1999) . The one thing both sides seem to agree on is that reef ecosystems as we know them, will not survive without immediate action to prevent further anthropogenic stress and exploitation (Pandolfi et al. 2003).

While coral diseases are occurring globally, their incidence appears to be the most severe in the Caribbean (Porter and Meier 1992; Grigg 1994; Hubbard et al. 1994; Hughes 1994; Jackson 1997,2001; Cooney et al. 2002; Bruno et al. 2003; Sutherland et al. 2004; Wapnick et al. 2004; Aronson and Precht 2006). Over the past few decades reports show that disease is responsible for a roughly 80% loss in Caribbean coral cover (Gardner et al. 2003; Nowak 2004;

Wapnick et al. 2004). The *Acropora* coral genus appears to have been the hardest hit by disease, with *A. palmata* showing a 90-95% decline across the Caribbean (Garrison et al. 2003) and *A. cervicornis* populations collapsing across the region (Wapnick et al. 2004). These extraordinary declines are of extreme importance to the health of the ecosystem, as until recently they served as the primary reef building corals for the area, responsible for developing reef framework (Lirman 1999), stabilizing the substrate and decreasing the impact of wave induces coastal erosion, providing crucial habitat to diverse reef fish populations (Lirman 1999), as well as other important reef organisms (Precht et al. 2002). As a result of these drastic declines the two *Acropora* species have been the first corals to ever be listed as Threatened Species under the Endangered Species Act (Precht et al. 2002; NMFS 2006). In the absence of the acroporids, *Gorgonia* and *Montastraea* species have become especially important to Caribbean systems in their attempts to maintain the reef framework and prevent the onslaught of macroalgae (Aronson and Precht 2001; Kim and Harvell 2002; Bruno et al. 2003).

3.3.3 Types of Coral Disease

As was mentioned in the previous chapters, the term "disease" refers to any deviation from the *healthy* state of an organism (Dorland 2000; Singleton and Sainsbury 2006; Stedman 2006). In contrast, "health" is defined as an individual's ability to resist or adapt to various stresses, whether they are physical, chemical, biological, social, etc. (Meade and Earickson 2000). Thus, a coral would be considered *diseased* when the holobiont is no longer able to withstand or adapt to an environmental insult causing the coral's function to decline such that their ability to survive is in jeopardy. Whereas, a *healthy* coral would be one which was able to successfully adapt to the insult and whose survival is not at risk.

3.3.3.1 Infectious vs. Noninfectious Diseases

As stated by Work et al. (2008), "disease is the outcome of complex interactions between the host, causative agent(s), and the environment." *Resistance* to disease would therefore be the "natural or acquired ability" of the coral holobiont to "maintain its immunity to or to resist the effects of" antagonistic biotic and/or abiotic agents (Madl 2005). Conditions, in which the etiolic agent is a living (biotic) organism, are usually referred to as "infectious" because the disease-causing agent can be passed from the infected host to a new host, either with or without direct contact. Conversely, noninfectious diseases refer conditions in which the etiologic agent is not transmittable, often because the disease is the result of non-living (abiotic), environmental stressors. However, determining whether the initial diseased state was caused by an infectious or noninfectious agent can be difficult, as "biotic and abiotic diseases are often closely related" (Sutherland et al. 2004). For example, given that corals have evolved to thrive in a relatively narrow range of environmental conditions, its stands to reason that subtle changes in the surrounding environment have the potential to stress the coral and decrease its ability to resist disease (Dubos 1961; Rosenberg and Ben-Haim 2002; Sutherland et al. 2004; Williams and Miller 2005; Lesser et al. 2007). In addition to reducing disease resistance, abiotic stressors are also known to promote the growth, virulence, and/or transmission rate of biotic pathogens, and cause resident microbes to become pathogenic (Kushmaro et al. 1996,1998; Toren et al. 1998; Ben-Haim et al. 1999; Alker et al. 2001; Banin et al. 2001a,b; Israely et al. 2001; Kuta and Richardson 2002; Sutherland et al. 2004).

3.3.3.2 Coral Bleaching

> "Although coral biologists have not considered bleaching as a disease (Peters 1984; Hayes and Goreau 1998; Richardson 1998), coral bleaching precisely fits the definition of disease – a process resulting in tissue damage or alteration of function, producing visible physiological or microscopic symptoms."
> – Rosenberg and Ben-Haim (2002)

Thermal Bleaching

Corals generally thrive in waters very close (within 1 to 2˚C) to their narrow thermal range (Glynn 1993; Berkelmans and Willis 1999; Harvell et al. 2001; Coles and Brown 2003; Fabricius 2006; Harley et al. 2006; Marshall and Schuttenberg 2006; Veron et al. 2009), when this thermal limit is exceeded for long periods of time the coral becomes stressed. When water temperatures exceed the thermal limit of the coral, the zooxanthellae function poorly. The loss of the zooxanthellae as a result of thermal stress is referred to as "thermal bleaching," or more commonly as "coral bleaching" (Hubbard 1997; Crabbe 2008). The zooxanthellae have pigment associated with them, when they are expelled only the clear tissue overlaying the white calcareous skeleton of the coral remains (**Figure 3.5**) – giving it the appearance of being "bleached" (Kleppel et al. 1989; Lang et al. 1992; Glynn 1993; Kushmaro et al. 1997; McCreedy et al. 2006; Lesser 2007). Initially, it was unclear whether the zooxanthellae were leaving the coral or the corals were forcing the zooxanthellae to leave (Strychar et al. 2004a,b); however, more recent evidence indicates that the corals are expelling their zooxanthellae in order to go into a hybernative-like state until the temperatures return to levels within the coral's optimal range (Crabbe 2008). If the temperature returns to normal relatively quickly (around a week) the coral will allow healthy zooxanthellae in the water column to re-inhabit their surface layer, allowing the coral to recover (Lang et al. 1992). However, prolonged thermal stress (lasting several months), usually in the form of increased or decreased sea surface temperatures (SSTs), or extreme SSTs, can result in coral mortality (Lang et al. 1992).

Conflicting Theories over the "Primary" Cause of Coral Bleaching

There is currently a debate within the scientific community as to whether or not bacteria can also cause corals to bleach. Eugene Rosenberg and his colleagues argue that specific bacteria (*Vibrio coralliilyticus* and *V. shiloi*) can cause corals (specifically *Oculina patagonica*,

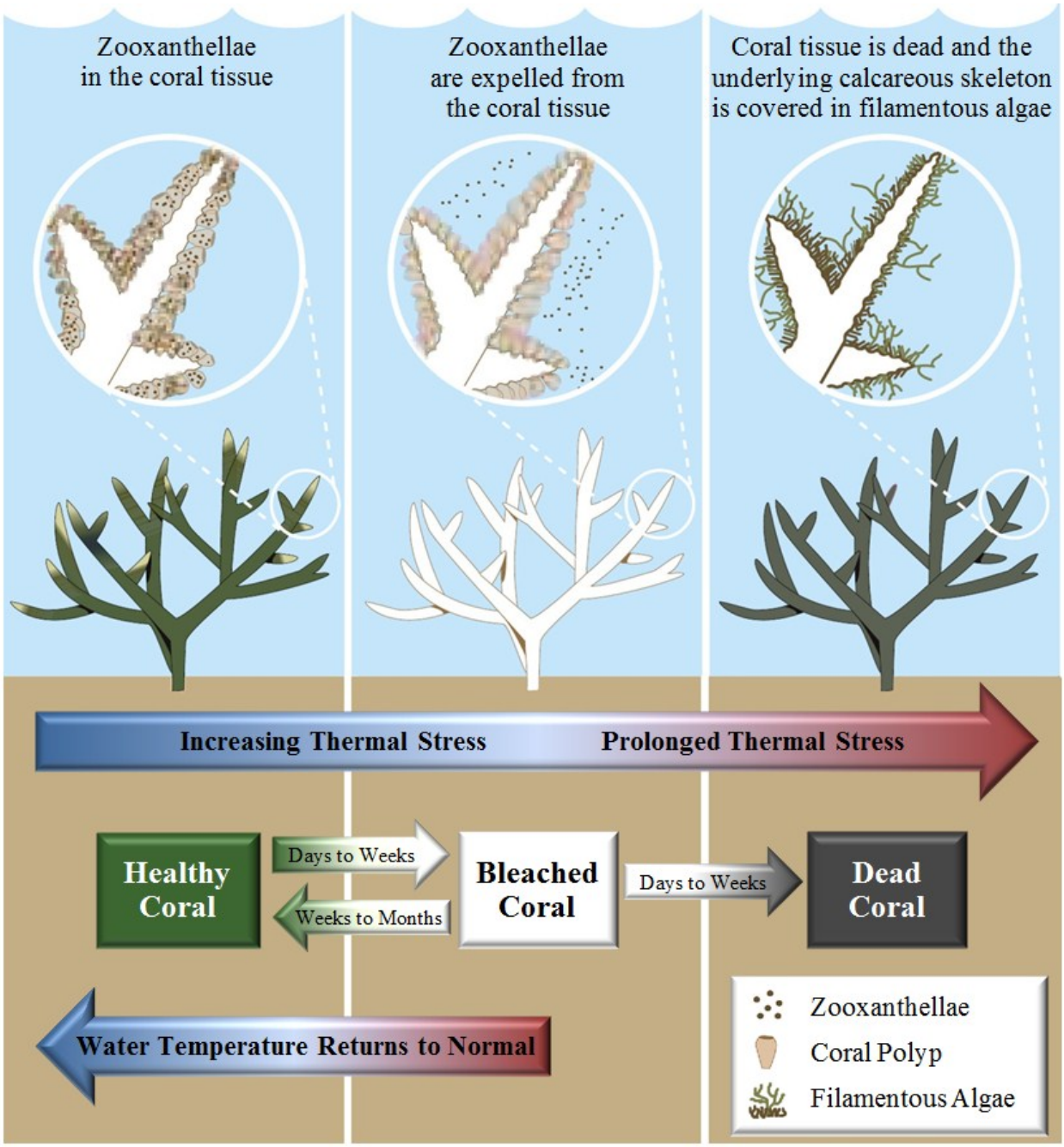


Figure 3.5 Diagram of thermal coral bleaching. Note: this diagram was adapted from Figure 1.3 on page 7 of Marshall and Schuttenberg (2006).

which is an invasive encrusting coral found in the Mediterranian) to become bleached (Kushmaro et al. 1996-1998,2001; Ben-Haim and Rosenberg 2002; Rosenberg and Ben-Haim 2002; Ben-Haim et al. 2003a,b; Rosenberg and Falkovitz 2004; Rosenberg et al. 2007a-b,2009b; Rosenberg and Kushmaro 2011). However, Ove Hoegh-Guldberg and his colleagues disagree, suggesting instead that abiotic, environmental stressors are the primary cause of the coral bleaching, and that if there is in fact any bacterial involvement having to do with the disease that is an opportunistic infection made possible by the weakened defenses of the coral due to the bleaching (Ainsworth et al. 2007a-b,2008a; Leggat et al. 2007a; Lesser et al. 2007; Ainsworth and Hoegh-Guldberg 2008,2009; Stat et al. 2009).

Rosenberg and his colleagues first started to promote their theory of "Bacterial Bleaching" in the late 1990s following two experiments by (Kushmaro et al. 1996,1998). They brought samples from the bleached lesion (the boundary between the bleached and non-bleached tissue) of *Oculina patagonica* back to their lab, where they were able to isolate *Vibrio shiloi* and obtain a pure culture of it. Koch's postulates were considered met when *V. shiloi was* re-isolated from the tissues of the experimentally bleached corals. They performed this, and variations of this experiment successfully on the same coral species between 1994 and 2002 (Kushmaro et al. 1996-1998,2001), during which time they identified *Hermodice carun-culata*, a marine fireworm, as the vector for keeping the pathogen alive during the winter months and then re-infecting the corals with *V. shiloi* the following summer.

They then used the results of these studies to develop their "Bacterial Bleaching Hypothesis" in which they proposed that bacteria was the primary driver of mass coral bleaching over the past 2 decades, rather than Climate Change (Rosenberg and Ben-Haim 2002). Then in 2003 they found that they were no longer able to isolate *V. shiloi* from the bleached corals, from which they concluded that the corals must have developed resistance to this bacterial pathogen

and adapted to the increased sea temperatures (Koren and Rosenberg 2006; Reshef et al. 2006; Rosenberg and Kushmaro 2011). This led them to come up with the "Coral Probiotic Hypothesis" which suggests that the coral holobiont is formed through the selection of the microbes best suited for a given coral host based on the current environmental conditions (Reshef et al. 2006; Rosenberg et al. 2007a,b). They then generalized this theory into the "Hologenome Theory of Evolution" (Rosenberg et al. 2007b,2009a,2010; Zilber-Rosenberg and Rosenberg 2008; Rosenberg and Zilber-Rosenberg 2011).

Hoegh-Guldberg and his colleagues were not necessarily opposed to the concept of coral bleaching being caused primarily by bacteria, but rather they questioned the research methods used. For example, Rosenberg's group performed all their experiments exclusively in aquaria (Ainsworth et al. 2008a; Ainsworth and Hoegh-Guldberg 2009), rather than in the ocean; which meant that there was no way of knowing whether or not their results were legitimate or just artifacts of the artificial environment in which the studies were performed. Additionally, all of their experiments were done on a coral known to be invasive to the area, and not representative of natural coral reef systems (Rosenberg and Falkovitz 2004).

In order to determine whether or not the results of Rosenberg's group were accurate, Hoegh-Guldberg and his colleagues performed a similar experiment on the same type of coral from the same geographic region; only this time they took their samples from corals in their natural environment during the annual summer bleaching period (Ainsworth et al. 2008a). Hoegh-Guldberg's group also took the sampling design one step further, by not only taking core samples from healthy *O. Patagonia* tissue and the bleaching lesion, but they also took a sampled of tissue that was entirely bleached (Ainsworth et al. 2008a). When they analyzed their results they found that while there was the same extent of bleaching as there had been in previous years, they found no signs of *V. shiloi* present in any of their samples, nor did they find the penetration

of and proliferation within the coral tissues by any type of *Vibrio,* or any other bacteria (Ainsworth et al. 2008a). In fact, they found no bacterial communities to be associated with any of their samples (bleached and non-bleached). The only microbial communities they encountered were endolithic communities found to occur under the bleached tissue as well as the pre-bleached and non-bleached coral tissues. During coral bleaching they found that there was a phase shift from the endolithic dominance, to cyanobacterial dominated groups just prior to bleaching, and then a final shift towards green algae during bleaching. From this Hoegh-Guldberg's group concluded that both the results and bleaching-related hypotheses of Rosenberg's group were incorrect (Leggat et al. 2007b; Ainsworth et al. 2008a).

Hoegh-Guldberg and his associates did suggest that the shift in microbial communities during bleaching might indicate that the coral was attempting to stabilize itself during a bleaching event, or it could also indicate potential sources for opportunistic pathogens (Ainsworth et al. 2008a). They went on to say that there was no doubt that the microbial communities play an important role in diseases (including bleaching), in fact that's obvious given the symbiotic nature of the holobiont. However, they warned that before experimental "solutions" such as the use of bacteriophages or the application of antibiotics were attempted, more research should be done on the potential presence of opportunistic pathogens, and how these treatments might affect them (Ainsworth et al. 2008a).

In a later study, Hoegh-Guldberg's group ran this experiment both in lab aquaria, to simulate Rosenberg's experimental design, and in the ocean, representing their own design and the natural conditions of the coral (Ainsworth and Hoegh-Guldberg 2009). The results of their studied showed that while the "penetration" into and "proliferation" of the bacteria within the outer layers of the coral tissues did occur in the aquaria, it did not occur in the ocean. This suggests that Rosenberg's results were largely an artifact of their research methods, rather than

of the actual conditions of the coral. They also cited a study by Kline et al. (2006) which found "major increases" in the abundance of bacteria just by maintaining corals in aquaria, without any type of experimental stressor (i.e. changing the temperatures, adding pathogens, etc.).

3.3.4 Possible Causes of Coral Diseases

As was shown in the conflicting theories regarding the cause of coral bleaching, there is currently much debate among the scientific community as to whether most coral diseases are caused primarily by infectious biotic pathogens or environmental stress, which in turn makes the coral more vulnerable to opportunistic secondary infections by biotic organisms. Consequently, while "identification of the causal agent(s) is typically the first question asked following an outbreak, it remains one of the most elusive aspects of coral disease epizootiology" (Williams and Miller 2005). Identifying the primary cause for specific coral diseases has been especially challenging given that individual corals with a given disease have been found to have different, potentially pathogenic, microbes present (Casas et al. 2004; Pantos and Bythell 2006; Thompson et al. 2006; Lesser et al. 2007; Toledo-Hernández et al. 2008; Sunagawa et al. 2009; Weil and Rogers 2011). Increasingly, studies are proposing that many, if not the majority, of coral diseases are caused by the infection of one or more opportunistic pathogens (Foley et al. 2005; Selig et al. 2006; Lesser et al. 2007; Weil and Rogers 2011), rather than a "single highly virulent primary pathogen" (Foley et al. 2005). This theory is supported by recent experiments, which suggest that external, environmental stressors can trigger changes in coral's microbial communities (Selig et al. 2006; Bally and Garrabou 2007; Abrego et al. 2008; Bourne et al. 2008; Bourne et al. 2009).

Rosenberg et. al.'s "Coral Probiotic Hypothesis" and "Hologenome Theory of Evolution" suggest that coral holobiont responds to stress by changing (or reconfiguring) the microbes present in order to increase their resistance to disease (Rosenberg et al. 2007b,2009a,2010;

Zilber-Rosenberg and Rosenberg 2008; Rosenberg and Zilber-Rosenberg 2011). However, while the microbial communities associated with healthy corals do appear defense mechanisms, such as the production of antimicrobials (Mullen et al. 2004; Sutherland et al. 2004; Kelman et al. 2006; Ritchie 2006; Lesser et al. 2007; Ward et al. 2007); recent studies show that environmental stress tends to compromise these defense systems (Ritchie 2006; Klaus et al. 2007; Lesser et al. 2007; Carilli et al. 2009; Carilli et al. 2010; Haapkylä et al. 2011) rather than activating or strengthening them. Given that most corals appear to become diseased following physiological stress caused by various types of "environmental insults" (Bally and Garrabou 2007; Lesser et al. 2007; Carilli et al. 2009; Sokolow 2009; Weil and Rogers 2011), it is important to understand the abiotic, environmental factors known to stress corals.

3.3.5 Coral Stressors

There are a number of stressors, which can cause decreased coral functioning, sterility, and mortality (Sammarco 1982; Bally and Garrabou 2007; Anthony et al. 2008; Anthony et al. 2009; Harvell et al. 2009; Miloslavich et al. 2010). The major stressors include: ocean temperatures outside ideal thermal range, ocean acidification caused by increased carbon dioxide levels, large changes in salinity and nutrient levels, increased exposure to ultra-violet radiation, general pollution, physical damage, sedimentation, invasive species, over-fishing, and disease (Buddemeier et al. 2004; Anthony et al. 2007; Bally and Garrabou 2007; Donner and Potere 2007; Lesser 2007; Lesser et al. 2007; Sokolow 2009; Haapkylä et al. 2011). Of these, bleaching and other diseases have been the most widespread and severe (Ainsworth et al. 2008b), resulting in massive coral mortalities and ultimately community shifts towards algal species and away from reef-building corals (McClanahan and Muthiga 1998; Aronson and Precht 2000,2001; McClanahan et al. 2001). Recent studies of the Great Barrier Reef area show a marked decrease in both coral cover and diversity combined with an increase in seasonal disease outbreaks, which

are believed to be caused by a combination of agricultural runoff and climate change (Nowak 2004; Haapkylä et al. 2011).

Current climate change trends directly affect the marine environment by increasing the water temperature (Sokolow 2009; Lesser 2011; Richmond and Wolanski 2011; Voolstra et al. 2011), decreasing the pH (Hoegh-Guldberg et al. 2007; Sokolow 2009; Wooldridge and Done 2009; Hoegh-Guldberg 2011; Lesser 2011; Richmond and Wolanski 2011), causing sea level rise (Sokolow 2009; Hoegh-Guldberg 2011; Lesser 2011), changing ocean circulation patterns (Wilson et al. 2010; Lesser 2011), and decreasing salinity concentrations (Sokolow 2009; Haapkylä et al. 2011). Warming of tropical waters is believed to affect the phase of the North Atlantic Oscillation, which in turn causes increases in marine disease outbreaks throughout the Caribbean (Hayes et al. 2001; Hoerling et al. 2001; Harvell et al. 2002; Rosenberg and Ben-Haim 2002). Climate change, specifically climatic warming, is known to cause increased rates of pathogen development, survival, transmission, and host susceptibility, which together will cause increased levels of the incidence, diversity, and severity of coral diseases (Colwell 1996; Porter et al. 2001; Aronson et al. 2002; Harvell et al. 2002,2007; Weil and Croquer 2009). Increased water temperatures are also known to cause large-scale bleaching events, such as the 1998 El Nino global bleaching event which has been noted as the most severe and extensive bleaching occurrence on record (Harvell et al. 2002). Bleaching is also known to severely stress corals, making them much more vulnerable to other diseases (Glynn 1993; Brown 1997; Hayes et al. 2001; Rosenberg and Ben-Haim 2002; Selig et al. 2006).

This being said, there is the consistent and ever present problem of the lack of baseline data (Harvell et al. 2002). Further, it has been exceptionally difficult for scientists to understand the link between infectious disease and climate change, since associations between the two do not necessarily imply causation (Marcogliese 2001; Harvell et al. 2002). With the ever

increasing human population and globalization of the world market, comes severe and widespread, long-lasting environmental deterioration. Humans play a direct role in the transport of both marine species and marine pathogens (Carlton and Geller 1993; Harvell et al. 1999), in much the same way *Vibrio cholera,* a bacterial pathogen, was transported across the world in the ballast waters of oil tankers (Ben-Haim and Rosenberg 2002). Many scientists believe that anthropogenic stressors have stressed corals and their surrounding ecosystem, to such an extent that their resistance levels are severely compromised, while at the same time fostering the growth and virulence of their marine pathogens (Hayes et al. 2001; Harvell et al. 2002; Bruno et al. 2003,2007; Lesser et al. 2007; Selkoe et al. 2009; Pittman and Brown 2011).

3.3.5.1 Acidification

Over the last 4 billion years the chemistry of the world's oceans have undergone dramatic changes (Lunine 1999; Hamblin and Christiansen 2001; Garrison 2007). The majority of these changes were the result of various natural environmental processes that occurred during the formation of the oceans and earth's atmosphere. However, changes in the seawater chemistry of modern oceans are caused primarily by changes in the atmospheric concentrations of carbon dioxide (CO_2), nitrogen, and sulfur due to anthropogenic actions (Doney et al. 2007; Riegl et al. 2009; Erez et al. 2011). Until relatively recently, the average concentration of CO_2 in the atmosphere ranged from 200 to 300 parts per million (*ppm*; Fabricius 2008; Riegl et al. 2009). Atmospheric CO_2 concentrations began to increase dramatically (nearly 100 times faster than they had over the past 650,000) with the start of the Industrial Revolution in 18^{th} and 19^{th} centuries due to an exponential increase in the amount of fossil fuels being combusted (Siegenthaler et al. 2005; Fabricius 2008; Riegl et al. 2009; Schmidt and Wolfe 2009; Veron et al. 2009; Erez et al. 2011). Recent models are showing that the concentration of CO_2 in the atmosphere may double or even triple during the next century, increasing from the present

concentration of ~387*ppm* to 540-970*ppm* by the year 2100 (Siegenthaler et al. 2005; Hoegh-Guldberg et al. 2007; Fabricius 2008; De'ath et al. 2009; Riegl et al. 2009; Veron et al. 2009).

Changes in the concentration of CO_2 in the atmosphere are important to the survival of corals for several reasons. The first is that 20-50% of the CO_2 released during the burning of fossil fuels is absorbed by the ocean (Erez et al. 2011), which makes CO_2 one of the major agents regulating the buffering capacity of the marine environment (Erez et al. 2011). Typically, the pH of the ocean is considered to be slightly basic, oscillating between a pH of 7.7 and 8.3 (Madl 2005; Sokolow 2009; Erez et al. 2011). As CO_2 dissolves in the ocean it forms carbonic acid, which in turn lowers the pH of the ocean causing it to become slightly more acidic (Feely et al. 2004; Riegl et al. 2009; Sokolow 2009; Veron et al. 2009; Erez et al. 2011). Consequently, large increases in the amount of CO_2 being absorbed by the ocean results in acidification of the marine environment (Riegl et al. 2009; Sokolow 2009).

Given that the skeleton of hermatypic corals is made of calcium carbonate ($CaCO_3$), the pH of the ocean is especially important to corals. Decreases in the pH of the ocean make it more difficult for corals to secrete $CaCO_3$, which, in turn, slows the rate of growth (Kleypas et al. 1999a,b; Ohde and Hossain 2004; Orr et al. 2005; Guinotte and Fabry 2008; Riegl et al. 2009; Sokolow 2009). Additionally, if pH levels drop to (or below) 7.4, calcareous skeletons will not only stop growing, but they will begin to dissolve (Fine and Tchernov 2007b,a; Erez et al. 2011). As a result, ocean acidification is currently not only stressing corals and retarding their growth, but it has the very real potential of eliminating them all together (Veron 2008; Riegl et al. 2009; Erez et al. 2011). Recent studies show that if atmospheric CO_2 concentrations continue to increase as predicted, reefs world-wide will being to show signs of increased stress due to acidification in the next 20-40 years (Caldeira and Wickett 2003; Cao et al. 2007; Hoegh-Guldberg et al. 2007; Veron et al. 2009; Erez et al. 2011).

3.3.5.2 Thermal Stress

Corals must live in a specific environmental niche in order to satisfy the requirements of both their animal and plant-like functions (Buddemeier et al. 2004). Optimal living conditions for most corals require clear, shallow, oligotrophic waters within a relatively narrow thermal range. While some corals have been found to survive in temperatures as low as 16˚C (Dana 1843; Mayor 1915; Walton Smith 1971; Walker et al. 1982; Coles and Fadlallah 1991) and as high as 36˚C (Kinsman 1964; Walton Smith 1971; Jaap 1979; Walker et al. 1982; Coles and Fadlallah 1991; Hubbard 1997), the optimal thermal range for most tropical, shallow-water corals is between 25˚ and 29˚C (Vaughan and Wells 1943; Kinsman 1964; Jaap 1979; Walker et al. 1982; Hubbard 1997). Within this range corals will tend to grow faster or slower depending on the temperature (Mayor 1914,1915; Vaughan and Wells 1943; Hubbard 1997). Because corals have such a narrow optimal thermal range, relatively small fluctuations above or below this range can result in thermal stress (Walker et al. 1982; Hubbard 1997; Hoegh-Guldberg and Fine 2004). The severity of the thermal stress caused by anomalous sea temperatures depends on a number of factors, such as: how much deviation there is between the anomalous temperature and the optimal thermal range for that specific species of coral; how long the temperature anomaly lasts; and whether or not the coral experienced additional stress before, during, or after the change temperature (Hubbard 1997; Brown et al. 2002a,b; Carpenter et al. 2008; Middlebrook et al. 2008; Brandt 2009; Carilli et al. 2009, 2010).

Low levels of thermal stress are often exhibited as slowed growth rates, decreased reproductive success, and reduction in the coral's ability to defend itself from and/or withstand additional stressors (aggression by other corals, macroalgal overgrowth, fighting off diseases, etc). Moderate thermal stress often results in the suspension of all non-essential processes (such as growth and reproduction), and possible loss of their zooxanthellae (Walker et al. 1982;

Hubbard 1997; Crabbe 2008). High thermal stress often causes the corals of an entire reef to lose their zooxanthellae. Repeated, prolonged, or extreme thermal stress can result in coral mortality and possibly even the destruction of the entire reef system (Hubbard 1997).

The majority of the observed coral bleaching episodes, including the world-wide mass coral bleaching of 1998 and the Caribbean-wide mass bleaching of 2005 (reported as the worst bleaching event on record), have occurred during warm-water anomalies (Brown 1997; Glynn et al. 2001; Aronson et al. 2002a; Douglas 2003; Strychar et al. 2004a; Strychar et al. 2004b; Takahashi et al. 2008; Croquer and Weil 2009; Coffroth et al. 2010; Csaszar et al. 2010; Eakin et al. 2010). However, corals have also become bleached as the result of cold-water anomalies (Roberts et al. 1982; Walker et al. 1982; Hoegh-Guldberg and Fine 2004; Hoegh-Guldberg et al. 2005; Crabbe 2008; Lirman et al. 2011). Over the last few decades climate change has caused noticeable increases in the frequency, distribution, and severity of warm-water thermal bleaching events. It is important to note that it is not just the increasing temperatures we have to worry about, climate change models are also predicting more extreme weather indicating consecutive bleaching as the result of both high and low thermal anomalies is likely to become increasingly more common.

3.3.5.3 Sea-Level Rise and Changes in Carbonate Mineral Saturation

In order to understand how reefs will respond to both present and future changes in sea level, it is important to first understand how reefs evolved through geologic time. When the rates of sea level change, coral reefs must in turn adapt their accretion (growth) rates in order to survive. Reefs are generally found in warm, shallow waters, with the dominant reef-building corals (such as the *Acropora* spp.) generally located in waters less than 20 *meters* below the sea surface (see **Figure 3.1** on page 25). Below a "critical depth" of roughly 15*m* vital processes to the coral will begin to shut down (Schlager 1981 in Neuman and Macintyre 1985). In order to

remain above this critical depth when the rate of sea level rise increases, corals must adjust their growth rates to either match (for shallow reefs that are already near the surface) or exceed the new rate of sea level rise (for deeper reefs), or risk drowning (Hallock and Schlager 1986). Conversely, if the reef is near the sea surface and the rate of sea level rise abruptly slows down, then the reef must either decrease its growth rate or switch from vertical accretion to horizontal accretion moving either forward into the current (prograde) or backward (with the current). If the reef does neither it will continue to grow vertically until it is at or slightly above the sea surface, at which point a "capping" phase is often triggered, causing the top of the reef to be cemented over, preventing further growth.

This relationship between the rates of sea level rise and coral accretion controls both the internal and external makeup of reefs by creating complex internal mosaics and external geometrics characteristic of the changes the reef undergoes (Neumann and Macintyre 1985). Thus, examination of both these interior and exterior characteristics can facilitate the study of the relationship between corals and changing sea levels (Neumann and Macintyre 1985). By examining the succession in the vertical internal facies of cores that have been drilled from Holocene reefs, scientists are able to deduce the local sea level histories. Not long after scientists first started examining these reef cores that Neumann and Macintyre (1985) proposed that when faced with sea level rise, reefs had three options (**Figure 3.6**): they must either "keep-up," attempt to "catch-up," or ultimately "give-up" (Neumann and Macintyre 1985; Hubbard 1997).

As depicted in **Figure 3.6**, *keep-up* and *catch-up* reefs often alternate between being an *A. palmata* or *A. cervicornis* – dominated reef. Both types of *Acropora* are known to have rapid growth rates compared to other types of corals, making it well suited for attempting to *keep-up* or even *catch-up* to rising sea levels. *A. palmata* has a more narrow depth range, preferring only shallow waters; which is why it is often associated with *keep-up* reefs. Whereas, *A. cervicornis*

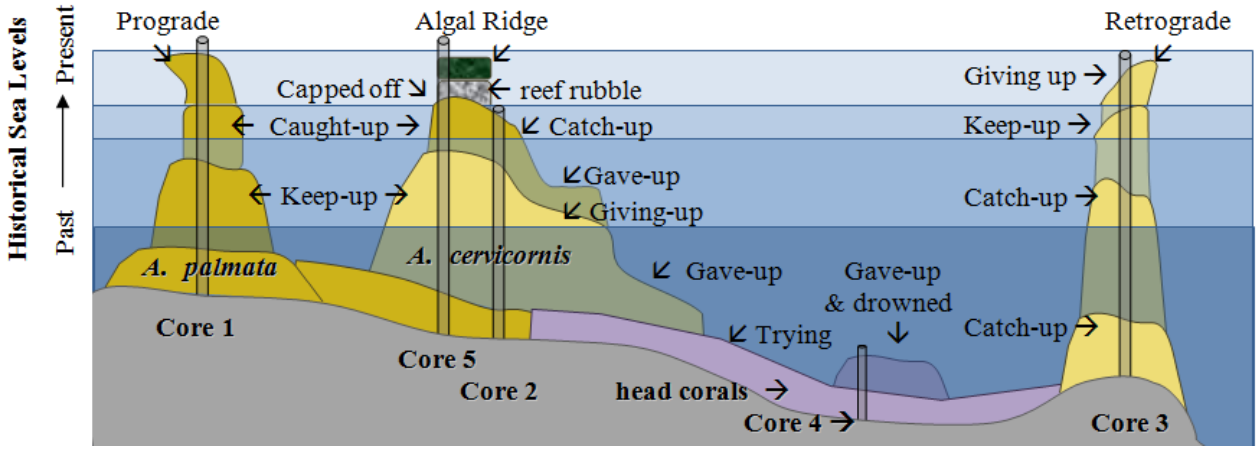


Figure 3.6 Diagram of the potential responses of various types of reefs to sea-level rise.

has a larger optimal depth range and can withstand and thrive in deeper waters, which is why it is often associated with *catch-up* reefs. Now what would happen if the coral was not able to increase its accretion rate enough, or even stopped accreting all together? In this case the coral would eventually, if it has not already, "give-up" and drown. This is often the case when deep water reefs, as they tend to be dominated by various types of slow growing head corals (depicted as the purple reef formations in **Figure 3.6**) which are not capable of achieving the rate of growth necessary to remain above their critical depth.

It should also be noted that head corals are not the only corals known to *give up* in the facing of rising sea levels. There are a number of factors that influence the relationship between rates of sea level rise and coral accretion; such as: how fast the rate of sea level rise is, whether or not the increase in sea level is smooth or oscillates, and whether or not the increase is constant or tends to abruptly start and stop (Neumann and Macintyre 1985). The geomorphology of the base of the reef, referred to as "antecedent topography," can also influence the reefs ability to adapt to the changing sea level. For example, reefs that have a sloped base often have prolonged start-up times, which means that once their *catch-up* phase does start, they will have further to go than they would have if their base had been more level. Other stressors such as changes in

temperature, salinity, turbidity, pollution, etc. can all stress the reef to such an extent that its accretion rates are slowed down dramatically, possibly even to the point at which the coral stops growing all together. If the reef is bordered by a shallow lagoon, there is an even greater chance of coming into contact with stressors like these.

3.3.5.4 Agricultural Run-off

Studies show that over the past century the humans have severely altered the global nitrogen cycle, causing a massive increase in the amount of nitrogen, as well as other nutrients, being deposited into the oceans (Vitousek 1994; Nixon 1995; Bruno et al. 2003). Agricultural runoff has been the primary source of the massive nutrient increase in the marine environment, followed by deforestation resultant runoffs, seepage from injection-well effluents through the water table, as well as some naturally caused nutrient increases such as: local upwellings and internal tidal bores (Shinn et al. 1994; Nixon 1995; Szmant 2002; Bruno et al. 2003; Leichter et al. 2003). While the degree to which the "nutrient effect" is happening and influencing different parts of the world it is not entirely understood. The one thing that is known is that this effect does exist and it poses a huge risk to the health and overall survival of marine systems as we know them because nutrients foster the spread of infections through feeding the microbial agents which make up the pathogens and stimulate algal growth (communication with John Bruno of the University of North Carolina cited in Nowak 2004). A study by Bruno et al. (2003) directly tested the impact of nutrients on two diseases, and found significantly positive correlation between nutrient enrichment and both yellow-band disease, which mainly effects reef building massive corals, and Aspergillosis, which targets sea fan corals.

Critics argue that any increases in nutrient levels would be "far too dilute" to adversely affect the corals (communication with Peter Ridd of James Cook University in Townsville, Australia cited in Nowak 2004). Bruno agrees that there is no data to support a Caribbean wide

increase in nutrient concentrations, however given the positive correlation between specific diseases and nutrients it is more likely that "local nutrient enrichment" is adversely affecting these marine systems on small, local scales (Bruno et al. 2003). The main fear is that if the degenerative changes in the Caribbean over the past 30 years are unprecedented in geologic history, than imagine how these environments will look in 100 years when the human caused stressors will be substantially greater (Bruno et al. 2003; Sutherland et al. 2004; Wapnick et al. 2004).

3.3.5.5 Aeolian Dust

During the latter part of the past century the volume of dust on the Earth's surface has risen dramatically due to wide scale droughts, poor resource management, agriculture resultant desertification, and natural cycles in general (Shinn et al. 2000; Pohl 2003). While global aeolian dust redistribution is a natural cycle, the newly formed increases in dust levels have made the impacts of this invasive dust more apparent and severe (Shinn et al. 2000; Pohl 2003). Some studies have found a direct correlation between Saharan dust and coral disease (Shinn et al. 2000; Harvell et al. 2002; Garrison et al. 2003).

3.3.5.6 Sewage and Other Pollutants

There is abundant evidence showing how severe pollution levels are adversely affecting the abundance, diversity, and overall habitat structure of coral reefs (Lewis 1984; Rogers 1990; Hughes 1994; Jackson 1997; Jackson et al. 2001; Pandolfi et al. 2003; Kaczmarsky et al. 2005; Klaus et al. 2007). In 2002, Katherine Patterson and her colleagues reported the first direct link between human sewage and coral disease (Patterson et al. 2002). Her study of the etiology of white pox disease revealed the first case of a marine invertebrate pathogen that was caused by a bacterial infection stemming from the human gut, the common fecal enterobacterium, *Serratia marcesens* (Patterson et al. 2002).

3.3.6 Effect of Disease on Dominant Reef-building Corals

When studying coral disease, or coral reefs in general, there is one genus which is almost always mentioned, *Acropora* (Scleractinia, Acroporidae). In addition to being highly speciose – with more than 100 known species (Wallace 1999; Veron 2000; van Oppen et al. 2001; Riegl 2002), and having one of the largest distributions – spanning globally, acroporids are also primary reef-builders worldwide (van Oppen et al. 2001; Precht et al. 2002). The genus is thought to date back to the Paleocene, providing a long geological and geo-spatial history of its species (van Oppen et al. 2001). In addition to providing habitat for many reef organisms, acroporids are most noted for their rapid rates of growth, reef accretion, and framework development (Precht et al. 2002).

Before diseases, acroporids were known for their ability to "grow like weeds," rebounding so rapidly from hurricanes that within a few years storm damage was almost indiscernible (Shinn 2004). A small, healthy colony of *Acropora cervicornis* had been known to grow an average of 10*cm* per year, expanding from a colony of 10 branches to over 59,000*m* of branches in only 10 years (Shinn 1976,2004). This genus has served as a foundation for healthy reefs around the world for thousands of years (Lewis 1984). In the last few decades, acroporids have taken a dramatic turn for the worse both on a regional and global scale. *Acropora* are generally the first species to experience severe declines during periods of disease epizootics and high stress (Riegl 2002).

In the last few decades, two coral diseases specific to *Acropora* (White-Band Disease and White Pox Disease) caused a dramatic decline in the two endemic Caribbean *Acropora* species, *A. palmata* and *A. cervicornis* (**Figure 3.7**). Both were candidates for Endangered Species status for 14 years (Richardson 1998; Green and Bruckner 2000; Patterson et al. 2002; Precht et al. 2002,2004); attaining official threatened status on June 8, 2006 (NMFS 2006). Similar trends

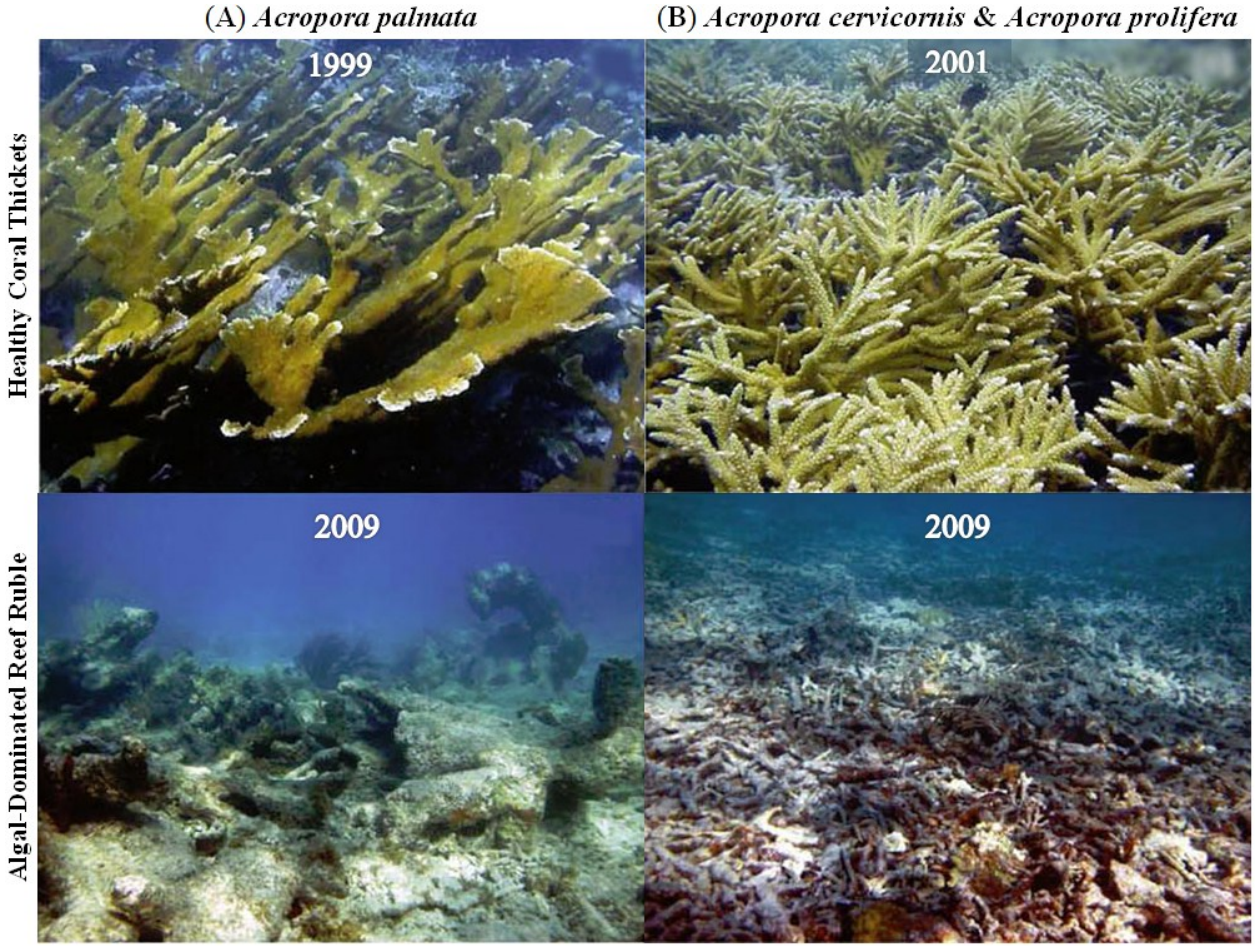


Figure 3.7 Time Series photographs depicting Caribbean *Acropora* species transitioning from healthy corals to algal-dominated, reef ruble in San Cristobal, Puerto Rico. (**A**) *Acropora palmata* in 1999 and 2009. (**B**) *Acropora cervicornis* and *Acropora prolifera* in 2001 and 2009. Note: the above figure was adapted from Figure 2 on page 555 of Bourne et al. (2009) with photographs taken by Ernesto Weil.

are following in the Indo-Pacific where, repeatedly, this genus was found to be the first infected and experience the highest terminal losses from density-dependent diseases (Riegl 2002). One study from the Arabian Gulf reported an *Acropora* mass mortality in the late 1990s, resulting in a decrease from ~ 80% total coral cover in 1996, to 0% in 1998, with no signs of substantial regrowth in following years (Riegl 2002). Riegl reports that, in addition to diseases occurring predominantly in *Acropora*, the diseases would kill all species within the genus and then either disappear along with the coral or move on to other, less susceptible hosts.

Theories regarding why *Acropora* species are so susceptible to disease include the following: (1) diseases preferentially target the most dominant, dense species (Riegl 2002; McClanahan et al. 2004a); (2) some species grow much slower when damaged (Winkler et al. 2004), thus reducing their ability to rebound from the pathogenic encroachment; (3) fragmentation from storms may increase susceptibility to infection (Winkler et al. 2004); (4) once damaged or infected the species exhibit marked decreases in reproduction, often ceasing all sexual forms of spawning all together; and last (5) stressed scleractinians, in general, have been shown to be more susceptible to pathogenic syndromes (Winkler et al. 2004). It is surprising that, while several studies suggest that acroporids are indicator species of reef decline, there is strong resistance from some corners to putting the two Caribbean species on the endangered species list (CoRIS 2004; Shinn 2004). **Figure 3.8** depicts the major diseases known to affect the *Acropora* coral genus world-wide.

3.4 Review of Current Research Methods

Although the field of coral pathology has received much attention lately, relatively little is known regarding the identification and understanding of causative agents of these diseases (Sutherland et al. 2004; Abrego et al. 2008; Bourne et al. 2008; Bourne et al. 2009; Correa et al. 2009; Mydlarz et al. 2010), especially when compared to the pathology of humans and other terrestrial organisms. As has been discussed in previous sections, progress in this area has been difficult due to the complex nature of both the coral holobiont and the surrounding marine environment (Ainsworth et al. 2008c; Ainsworth and Hoegh-Guldberg 2009). Recent technological advances have not only facilitated and improved our ability to study corals, but also greatly enhanced our understanding of coral histology (Bourne et al. 2009; Correa et al. 2009; Krediet et al. 2009; DeSalvo et al. 2010; Lins-de-Barros et al. 2010; Mydlarz et al. 2010; Kvennefors et al. 2011).

Figure 3.8 (shown on the following page) Diseases known to affect the *Acropora* coral genus worldwide. The diseases shown on the top row (**A** – **C**) are known to only *Acropora* corals world-wide; while the diseases on the second row (**D** – **E**) appear to only affect Caribbean *Acropora* species, and the diseases on the bottom two rows (**F** – **K**) affect *Acropora* in the Indo-Pacific. (**A**) Thermal Bleaching on *A. millepora* at the Great Barrier Reef, Australia. (**B**) Growth Anomalies (GA) on branching *Acropora* at the Great Barrier Reef, Australia. (**C**) Skeletal Anomaly (SKA) on *A. palmata.* (**D**) The two types of White-Band Disease (WBD), both depicted on *A. palmata*; WBD Type I is depicted in $\mathbf{D_1}$ and WBD Type II is depicted in $\mathbf{D_2}$. (**E**) White-Pox Disease (WPD or WPox), also known as Acroporid Serratiosis (APS), on *A. palmata*. (**F**) Black-Band Disease (BBD) on an *Acropora* species. (**G**) Brown-Band Disease on a branching *Acropora* species at the Great Barrier Reef, Australia. (**H**) Skeleton Eroding Band Disease (SEB) on *A. intermedia* at the Great Barrier Reef, Australia. (**I**) White Syndrome (WS) on a plating *Acropora* species at the Great Barrier Reef, Australia. (**J**) Yellow-Band Disease (YBD) on *A. pharaonis*. (**K**) BBD and SEB on the same colony of *A. muricata* at the Great Barrier Reef, Australia. Note: the above disease depictions were taken from the following sources: (**A**) taken from Figure 2 on page 1361 of Jones et al. (2008); (**B**,**G**,**H, I**) photos were taken by Betty Willis and published in Figure 8 on page 183 of Harvell et al. (2007); (**C** and $\mathbf{D_1}$) taken from Figure 3 on page 282 of Sutherland et al. (2004); ($\mathbf{D_2}$, **E**) photos were taken by Ernesto Weil and published in Figure 3 on page 178 of Harvell et al. (2007); (**F**) taken from page 29 of Raymundo et al. (2008); (**J**) taken from page 22 of Korrubel and Riegl (1998); and (**K**) taken from Figure 4 on page 47 of Page and Willis (2006).

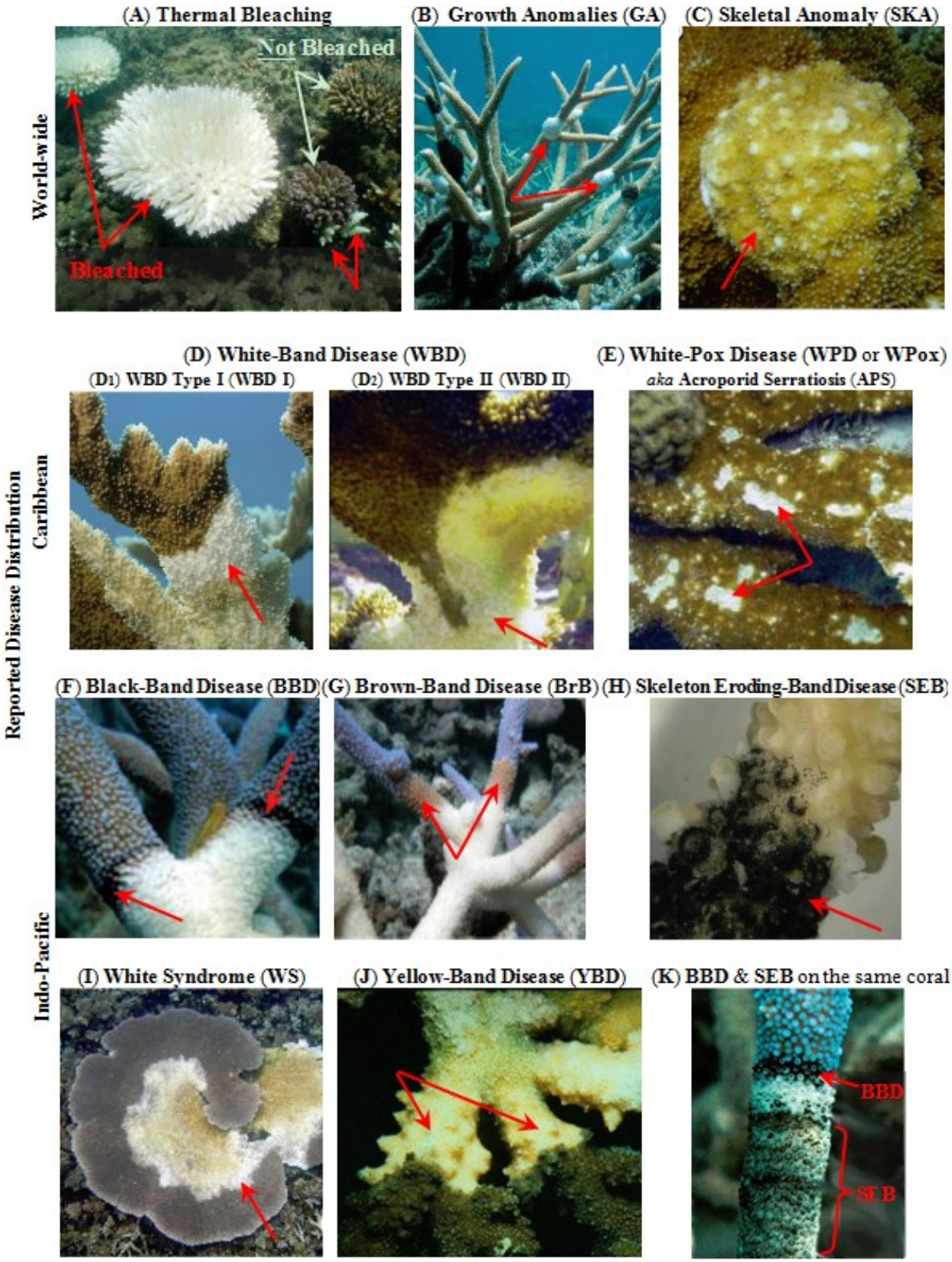


Figure 3.8 (see figure legend on the preceding page)

also greatly enhanced our understanding of coral histology (McClanahan et al. 2004b; Peters 2006; Ainsworth et al. 2007a; Abrego et al. 2008; Ainsworth and Hoegh-Guldberg 2008; Bourne et al. 2008; Bourne et al. 2009; Correa et al. 2009; Krediet et al. 2009; DeSalvo et al. 2010; Lins-de-Barros et al. 2010; Mydlarz et al. 2010; Kvennefors et al. 2011).

However, our overall understanding of coral pathology, specifically the *epizootiology*, *etiology*, and *histopathology* of coral diseases (see **Table 3.1** for definitions), has been hindered by the following problems with current research methods: (1) lack of consensus concerning the nomenclature associated with coral pathology; (2) how to correctly identify and distinguish between specific coral diseases in the field; and last, (3) the basis for etiologic diagnoses and overall design of current epidemiological models.

3.4.1 Nomenclature

The lack of a consensus on the terminology associated with coral diseases makes it difficult to differentiate between different diseases and for researchers to collaborate. In recent years there has been a push to use the medical community's standard disease-related terminology, such as the terms shown in **Table 3.1** (Sutherland et al. 2004.; Peters 2006; Work and Aeby 2006; Work et al. 2008). As part of this effort researchers from around the world have been working to create a common coral disease nomenclature system (Work and Aeby 2006; Raymundo et al. 2008; Woodley et al. 2008; Work et al. 2008; ICRI/UNEP-WCMC 2010a,b). However, despite these efforts there continues to be disagreements over the correct definitions and appropriate use of these terms. While most coral disease literature that provides a definition of the term "disease" uses some variation of the *standard medical definition*[1] in which a *disease* is essentially any condition that impairs the normal functioning of an organism (Peters 1997; McCallum et al. 2004; Rosenberg 2004; Sutherland et al. 2004; Weil 2004; Work and Aeby

[1] as defined by popular medical dictionaries, such as the Dorland's and Stedman's Medical Dictionary series

Table 3.1 Important Medical Terminology Related to Coral Diseases

Apoptosis	Programmed cell death
Atrophy	A wasting of tissues, organs, or the entire body, as from death & reabsorption of cells, diminished cellular proliferation, decreased cellular volume, pressure, ischemia, malnutrition, lessened function, or hormonal changes
Defense Mechanism	A physiological self-protecting response of an organism to a harmful stimuli
Epizootiology	The study of the distribution and determinants of health-related states or events in specified animal populations, and the application of this study to control health problems
	Note: The term "Epidemiology" should be used only when referring to human populations
Etiology	The science and study of the causes of disease and their mode of operation
Histology	The science concerned with the minute structure of cells, tissues, & organs in relation to their function
Histopathology	Science or study dealing with the cytologic & histologic structure of abnormal or diseased tissue
Hyperplasia	An increase in the number of normal cells in a tissue or organ, not due to tumor formation
Hypertrophy	General increase in bulk or a part of an organ, not due to tumor formation
Infectious	A disease capable of being transmitted from patient to patient, with or without actual contact
Lesion	A wound or injury; a pathologic change in the tissues. One of the individual points or patches of multifocal disease
Necrosis	Pathologic death of one or more cells, or of a portion of tissue or organ, resulting from irreversible damage
Neoplasia	The pathologic process that results in the formation & growth of a neoplasm (tumor)
Panzootic	An epizootic occurring on a global scale
Parasite	An organism that lives on or in another & draws its nourishment therefrom
Pathogen	Any virus, microorganism, or other substance causing disease
Opportunistic P.	an organism that is capable of causing disease only when the host's resistance is lowered
Pathology	The form of medical science & specialty practice concerned with <u>all</u> aspects of disease
Sign	Any abnormality indicative of disease, discoverable on examination of the patient; an <u>objective</u> indication of disease.
	Note: the term "symptom" refers to <u>subjective</u> indications of disease; consequently humans are the <u>only</u> type of animal that has *symptoms* associated with a given disease.
Stress	Reactions of the body to forces of a deleterious nature, infections, & various abnormal states that tend to disturb its normal physiologic equilibrium (homeostasis)
Stressor	An event or association that triggers a stress response
Susceptibility	Likelihood of an individual to develop ill effects from an external agent

Note: the above definitions were adapted slightly from the 28th Edition of Stedman's Medical Dictionary (2006)

2006; Raymundo et al. 2008; Woodley et al. 2008; Work et al. 2008) , many of these same researchers disagree as to which conditions *should* be classified as "diseased."

For example, one common dispute is over whether coral bleaching should be classified as a disease. Most researchers would agree that the normal functioning of corals becomes impaired during the bleaching process; which should mean that bleached corals are diseased regardless of whether their condition was caused by bacteria or thermal stress. The following studies agree with this logic, arguing that coral bleaching should be considered a type of coral disease (Rosenberg and Ben-Haim 2002; Jokiel 2004; Rosenberg 2004; Weil 2004; Rosenberg and Barash 2005; Woodley et al. 2008). However, other studies disagree, arguing that bleaching is not a disease because it is caused by physiological (namely thermal) stress (Hayes and Goreau 1998; Richardson 1998). The latter argument is supported by the vast majority of coral publications which refer to coral bleaching and coral disease separately through the use of phrases like "coral bleaching and disease" rather than "coral bleaching and other diseases", implying (intentionally or unintentionally) that bleaching is not a disease.

Another common dispute has to do with the difference between "diseases" and "syndromes," and how the standard medical definitions of the two terms (see **Table 3.2**) should be interpreted with regard to corals. Some argue that, based on these definitions, the two terms are synonymous and can be used interchangeably (Sutherland et al. 2004; Work and Aeby 2006); while others argue that the term "syndrome" should be used when referring to a poorly understood conditions, reserving "disease" for only those conditions in which the causative agent(s) have been identified (Hayes and Goreau 1998; Richardson 1998; Weil 2004; Lesser et al. 2007; Sheppard et al. 2009). Of the researchers in favor of distinguishing between *diseases* and *syndromes*, there is additional disagreement over what criteria must be met in order for something to be considered an etiologic agent.

Table 3.2 Comparing the *standard* medical definitions of "disease" and "syndrome" used by the most common medical dictionaries series both in print (Stedman's and Dorland's) and online (MedicineNet.com and MedlinePlus.com).

Medical Dictionary	Disease	Syndrome
Stedman (2006)	"1. An interruption, cessation, or disorder of a body, system, or organ structure or function. SYN illness, morbus, sickness. 2. A morbid entity ordinarily characterized by two or more of the following criteria: recognized etiologic agent(s), identifiable group of signs and symptoms, or consistent anatomic alterations. SEE ALSO syndrome." (page 550)	"*This word is not properly applied to a solitary symptom or sign.* The aggregate of symptoms and signs associated with any morbid process, together constituting the picture of the disease. SEE ALSO disease." (page 1888)
Dorland (1994)	"any deviation from or interruption of the normal structure of function of any part, organ, or system (or combination thereof) of the body that is manifested by a characteristic set of symptoms and signs and whose etiology, pathology, and prognosis may be known or unknown." (page 478)	"[a] set of symptoms which occur together" (page 1632)
MedicineNet (2011)	"Illness or sickness often characterized by typical patient problems (symptoms) and physical findings (signs)."	"A set of signs and symptoms that tend to occur together and which reflect the presence of a particular disease or an increased chance of developing a particular disease."
MedlinePlus (2003)	"An impairment of the normal state of the living animal or plant body or one of its parts that interrupts or modifies the performance of the vital functions, is typically manifested by distinguishing signs and symptoms, and is a response to environmental factors (as malnutrition, industrial hazards, or climate), to specific infective agents (as worms, bacteria, or viruses), to inherent defects of the organism (as genetic anomalies), or to combinations of these factors **: SICKNESS, ILLNESS** —called also *morbus*; compare **HEALTH"**	"A group of signs and symptoms that occur together and characterize a particular abnormality"

Some take a very strict approach, arguing that causation can only be *proven* through the fulfillment of Koch's postulates (Hayes and Goreau 1998; Richardson 1998). Adherence to this logic would mean that only those conditions caused by a single, biotic pathogen should be classified as coral diseases. As will be discussed in section 3.4.3, Koch's postulates are very limited, and not well suited for marine studies. Consequently, conditions caused by abiotic agents or otherwise brought about through physiologic stress would not qualify as coral diseases; nor would conditions caused by more than one biotic pathogen. Using this logic, not only would thermal bleaching and Black-Band Disease (which arguably have the most studied and understood etiologies) not be considered diseases, but some might argue that even those conditions that have fulfilled Koch's postulates were actually opportunistic infections following environmental stress and therefore would not fall under the category given that their primary cause was abiotic (Lesser et al. 2007).

While others take a slightly more relaxed approach, using the two terms to distinguish between conditions based on how much is known about their causative agent(s) and general etiology (Weil 2004; Lesser et al. 2007; Sheppard et al. 2009), rather than whether or not the condition is caused by an etiologic agent capable of meeting the requirements laid out by Koch's postulates. Universal adherence to this approach has the potential to facilitate comparisons of specific coral conditions between different studies. However, it is also problematic as most studies classify coral conditions based on the presence or absence of visually identifiable, gross (large) lesions or otherwise signs of *disease*, and are often unable to confirm their visual diagnosis through analyses performed at the microscopic level (Work and Aeby 2006; ICRI/UNEP-WCMC 2010a). For this reason, the approach laid out by Work and Aeby (2006) in the following section (3.4.2) would likely provide a more robust foundation for cross-study comparisons of different coral health, than would distinguishing between *diseases* and

syndromes while continuing to use vague names such as black-band, yellow-band, white-band, white-pox, white-plague, etc.

3.4.2 Identification

Currently, most *diseases* and/or *syndromes* are named according to the color and pattern changes they elicit in the tissues of the infected coral (e.g. white, black, red, or yellow band, line, blotch, spot, or pox disease). This system is problematic for many reasons. First, the resulting names are vague and subjective, which creates inherent confusion and difficulty distinguishing between the various *band*, *line*, *patch*, *blotch*, *pox*, *spots* diseases (Work and Aeby 2006; Ainsworth et al. 2007b). Additionally, corals are relatively simple organisms (especially when compared to other animals, such as humans), and as such, they have a limited number of different ways stress can be visibly expressed (i.e. "signs" of disease) by the coral at the gross level (Sutherland et al. 2004; Work and Aeby 2006; Weil and Rogers 2011). Consequently, the same visible signs could be caused by different diseases (Ainsworth et al. 2007b; ICRI/UNEP-WCMC 2010a); or multiple, concurrent diseases. Further, diseases that affect different types of corals, may illicit different signs based on the physiology of the coral it is infecting. Thus, one disease could be identified as several diseases, depending upon its stage of development/progression, the type of coral it is affecting, and whether other diseases and/or stressors are influencing the visual signs of the coral's compromised health.

To avoid the problems listed above, Work and Aeby (2006) suggest that when data are collected in the field, researchers should follow a standardized protocol for describing the lesions (signs of disease) on each coral, without trying to diagnosis what disease is causing the lesion.

> While there is no one right way to describe a lesion, the description should be explicit, concise, and provide applicable information on distribution, location on colony, edges, margins, shapes, relief, texture, color, size, and structures affected. Use of appropriate terminology aids brevity (Fig. 1, Tables 1 and 2). – Work and Aeby (2006)

Figure 3.9 on the following page summarizes the diagnostic protocol laid out in the figure and tables referred to in the above quote. Collecting data using the 15 criteria listed in **Figure 3.9** enables researchers to generate "concise and objective" morphologic descriptions of the disease lesions, which can then be used to generate strong case definitions of the observed disease (Work and Aeby 2006). Removing the subjective names and descriptions from coral disease data further facilitates cross-study comparisons, because the same terminology is used. This is especially important given that corals may exhibit the same external signs of stress as the result of different diseases or other environmental stressors.

3.4.3 Epidemiologic Models and Etiologic Diagnoses

The differences between terrestrial and marine systems make "the direct application of most epidemiological models difficult to interpret for marine systems" (Sokolow et al. 2009).

> For example, transmission rates in epidemic models typically depend on contact rates among susceptible and infectious hosts (McCallum et al. 2001), and this concept is not interpretable when hosts, such as coral, are sessile for most of their lives. Rather, movement and survival of pathogens outside hosts must be incorporated.
>
> – Sokolow et al. (2009)

Additionally, the fluid nature of the marine environment makes the implementation of standard epidemiological protocols difficult, if not impossible. One of the first measures taken when dealing with a disease outbreak, particularly those involving new diseases with unknown etiologies but appearing to be contagious, is to start isolation, containment, and quarantine procedures. However, in the case of coral epizootics, not only is isolation of a diseased coral impractical, but even if the coral could be quarantined through the use of some type of physical structure this would likely do more harm than good. Even though corals are considered to be relatively simple organisms (Weil and Rogers 2011), researchers are just starting to understand the complex nature of the holobiont (Reshef et al. 2006). Complete containment of a diseased coral would make the ocean water included in the containment device stagnant, potentially

Figure 3.9 (*shown on the following page*) Types of information that should be recorded when describing disease lesions on corals in the field. Note: this figure summarizes the information provided in Work and Aeby's (2006) Tables 1-2 and Figure 1 on pages 156-157. The depictions of coral types shown in **1**(**a-e**) are from page 485 of Veron and Wallace (1984) and are available online at: *http://biophysics.sbg.ac.at/coral/morfacro.htm*; the depiction of the "free-living" coral type shown in **1**(**f**) was taken from a Tiwan study that is available online at: *http://163.26.138.2/dyna/webs/index.php?account=admin&id=22&mod_area=15*; the coral image used for **2**(**a-f**), **3**(**a-d**), and **8**(**a-k**) was adapted from the Brain Coral depicted on page 87 of Humann and Deloach (2002); and last the images shown in **4-7** were taken from Figure 1 on page 157 of Work and Aeby (2006).

1. Describe what Type of Coral is the lesion found on

Identify the general colony type according to its overall structure & if possible specify the *Genus* and *species*

(**a**) Massive (**b**) Encrusting (**c**) Laminar/Explanate (**d**) Corymbose (**e**) Branching/Foliaceous/Arborescent (**f**) Free-Living

2. Describe Where the lesions are located on the colony

(**a**) Apical (**b**) Medial (**c**) Basal (**d**) Central (**e**) Peripheral (**f**) Colony-wide

3. Describe the how the lesions are Distributed across the surface of the colony

(**a**) Focal (**b**) Multifocal (**c**) Multifocal to Coalescing (**d**) Diffuse

4. Describe the appearance of the lesion Edges

(**a**) Distinct (**b**) Indistinct (**c**) Annular

5. Describe the appearance of the lesion Margins

(**a**) Smooth (**b**) Serrated (**c**) Undulating (**d**) Serpiginous

6. Describe the Shape of the lesions

(**a**) Circular (**b**) Oblong (**c**) Pyriform (**d**) Cruciform (**e**) Linear (**f**) Lanceolate (**g**) Irregular

7. Describe any three-dimensional structure associated with the lesion (i.e. the Relief of the lesion)

(**a**) Umbonate (**b**) Bosselated (**c**) Nodular (**d**) Exophytic (**e**) Fimbriated

8. Describe the Color of the lesion

(**a**) White (**b**) Black (**c**) Tan (**d**) Brown (**e**) Red (**f**) Orange (**g**) Yellow (**h**) Green (**i**) Blue (**j**) Purple (**k**) Pink

9. Describe the Size of the lesions

(**a**) Small (**b**) Medium (**c**) Large (**d**) Physical Measurement

10. Describe the Number of lesions on the colony

(**a**) Small (**b**) Medium (**c**) Large (**d**) Actual Count

11. Lesion Texture

(**a**) Rugose
(**b**) Smooth

12. Lesion Extent (% of Surface Area Covered)

(**a**) Mild (< 20%)
(**b**) Moderate (21-50%)
(**c**) Severe (> 50%)

13. Time (Rate of Lesion Onset)

(**a**) Acute (hours – days)
(**b**) Sub-Acut (weeks)
(**c**) Chronic (months – years)

14. Tentative Categorization of Lesion

(**a**) Tissue Loss (**b**) Discoloration (**c**) Growth Anomaly

15. Structures of the Coral Affected by the Lesion

(**a**) Polyp (**b**) Coenosarc (**c**) Skeleton

Figure 3.9 (see figure legend on the preceding page)

leading to eutrophic conditions, as well as starving both the symbiotic microbial inhabitants and their coral host of essential nutrients, and potentially causing the contained water to increase in temperature. Additionally, the device itself might shade the coral preventing the zooxanthellae from receiving the sunlight needed to undergo photosynthesis. Given that quarantine measures could potentially protect the surrounding healthy corals, the additional harm caused to the contained diseased coral might be considered an *acceptable risk*. However, this type of containment would only work for coral diseases caused exclusively by contagious biotic pathogens, transmitted through either the water column or biotic vectors (fish, snails, etc.). The primary type of diagnosing this type of disease is through the fulfillment of Koch's postulates.

3.4.3.1 Koch's Postulates

Of the numerous coral diseases that have been identified over the last few decades, Koch's postulates have only been fulfilled for the following five diseases (shown in **Figure 3.10**): White Plague II (Richardson et al. 1998b; Denner et al. 2003); White-Band II (Ritchie and Smith 1998); White Pox, also referred to as Acroporid Serratiosis (Patterson et al. 2002; Sutherland et al. 2010,2011); Aspergillosis (Smith et al. 1996,1998; Geiser et al. 1998); and Bacterial Bleaching (Kushmaro et al. 1996-1998,2001; Ben-Haim and Rosenberg 2002; Ben-Haim et al. 2003a,b).

While many scientists argue that fulfillment of Koch's postulates serves as definitive proof that the etiologic (causative) agent of the given coral disease was biotic in origin (Kushmaro et al. 1996-1997,2001; Richardson 1998; Ben-Haim and Rosenberg 2002; Patterson et al. 2002; Ben-Haim et al. 2003a; Denner et al. 2003); others argue that Koch's postulates should not be used as the standard as there are numerous diseases for which the postulates cannot or should not be fulfilled (Fredricks and Relman 1996; US EPA 2000; Banin et al. 2001a,b; Ritchie et al. 2001; Sutherland et al. 2004; Selig et al. 2006; Lesser et al. 2007; Work et al. 2008; Sokolow 2009). Additionally, Lesser et al. (2007) argue that most coral diseases, including those

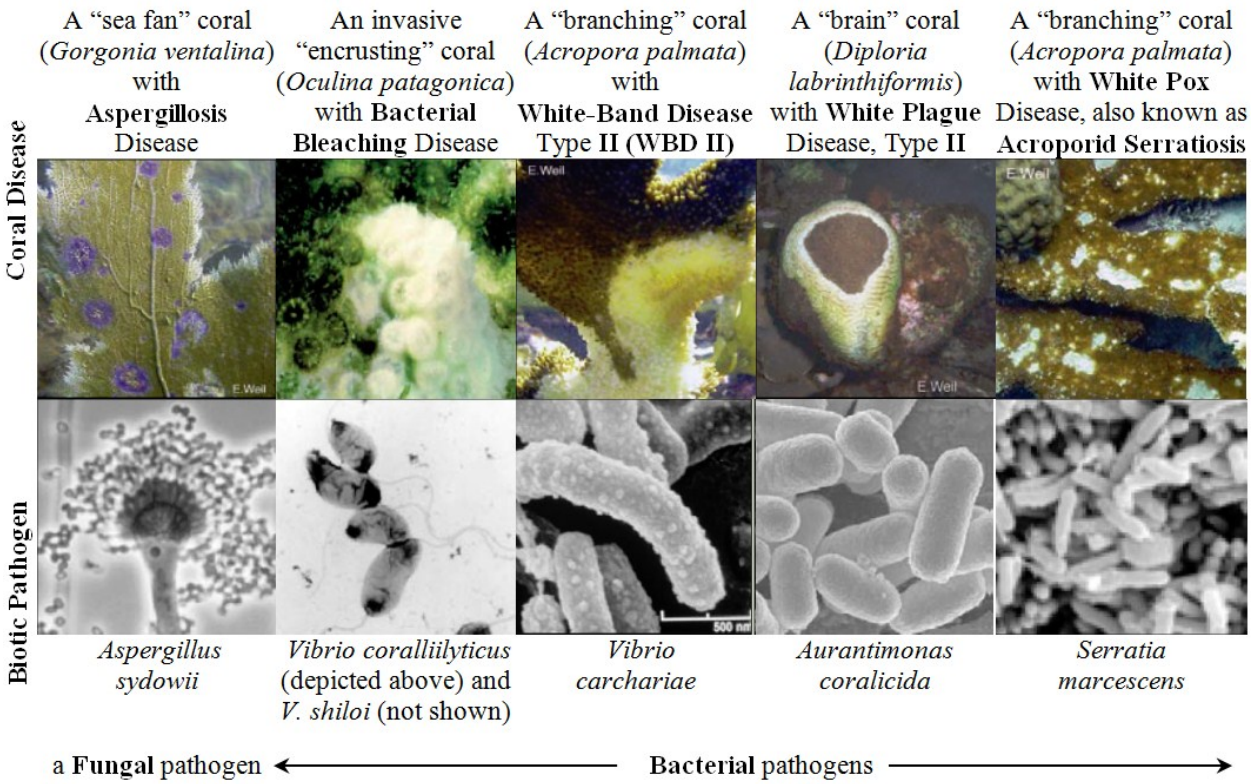


Figure 3.10 The five coral diseases in which Koch's postulates have been fulfilled, indicating that each is a biotically induced disease caused by the microbial pathogen indicated below. Note: this figure was adapted from Figure 3 on page 178 of Harvell et al. (2007).

for which Koch's postulates have been fulfilled, are actually secondary infections caused by opportunistic pathogens and that the primary cause of the disease is the physiological stress caused by abiotic environmental factors.

There are numerous potential problems with using Koch's Postulates in marine systems.

> Koch's postulates cannot be fulfilled according to the strict definition of the procedure for diseases that: (1) are caused by unculturable bacteria, fungi, or viruses (2) are caused by a consortium of microorganisms, (3) are caused by abiotic stressors, (4) require a vector or a carrier state, (5) cause subclinical or latent infection, or (6) cause injury through systemic attack via virulence factors such as toxins (Fredricks and Relman 1996; US EPA 2000). – Sutherland et al. (2004)

Additionally, the very nature of corals makes the postulate almost impossible to use, as (1) it is almost impossible to replicate "normal" physical and chemical reef environments in a laboratory;

(2) since the natural modes of pathogenic infection are not known it is impossible to simulate them through lab inoculations; and (3) as many presumed pathogenic microbes are found on the surface of corals, it is hard to transport them back to the lab without damaging the microbe or contaminating the aquarium water unintentionally (Sutherland et al. 2004; Selig et al. 2006). It is also important to note that the fulfillment of Koch's postulates, as well as other more modern molecular techniques, "do not prove disease causation" (Sutherland et al. 2004). However, they are important in that their results may indicate the presence of associations between potential biotic pathogens and a given disease, and thus furthering our understanding of the etiologic process for the given infectious disease (Ritchie et al. 2001; Sutherland et al. 2004).

Given the rapidly deteriorating condition of corals reefs worldwide, coupled with the grim outlook for their future, it is clear that substantial changes and progress needs to be made in the current methods being used to study these diseases. First, there needs to be more agreement among researchers as to the disease nomenclature used concerning corals. Second, more detail needs to be recorded when performing disease surveys in the field, and it is important that a standardized approach and specific terminology (such as that proposed by Work and Aeby 2006) be used. Third, current epidemiological models need to be adapted for the marine environment, including creating alternative criteria for disease causation for cases in which Koch's postulates are not appropriate (Sutherland et al. 2004). Last, as described in the following section (3.5), a geospatial analytical component needs to be added to these epidemiological models so that the spatial nature of these epizootics can be studied at local, regional, and global scales.

3.5 Geospatial Analysis and Coral Epizootiology

Increasingly studies are reporting that spatial distribution and patterns of both disease and other environmental factors may improve our understanding of both the cause and transmission dynamics of various coral diseases (Real and McElhany 1996; Foley et al. 2005; Crowder et al.

2006; Grober-Dunsmore et al. 2006; Ritchie 2006; Selig et al. 2006; Jones et al. 2008; Selkoe et al. 2009; Sokolow 2009; Weil and Croquer 2009; Zvuloni et al. 2009; Eakin et al. 2010; Ruiz-Moreno et al. 2010; Selig et al. 2010; Maina et al. 2011; Pittman and Brown 2011). As part of this recognition, the recording of GPS coordinates has become standard protocol for many coral field studies (Ginsburg 2000; Weil et al. 2002; Lang 2003; Willis et al. 2004; Grober-Dunsmore et al. 2006; Mayor et al. 2006; ICRI/UNEP-WCMC 2010b; Pittman and Brown 2011).

However, despite this growing recognition of the importance of studying the spatial nature of both coral diseases and their environmental stressors, only a handful of studies have actually used spatial statistics to analyze their spatial data (Jolles et al. 2002; Foley et al. 2005; Zvuloni et al. 2009). Further, the type of spatial analysis used by all of these studies was the *Ripley's K* statistic, which provides no spatial (mappable) output. Consequently, while these studies *were* able to determine whether or not diseased corals were spatially clustered (and, if so, at what spatial scales this clustering was occurring at), their results provided no indication of *where* these clusters were occurring.

Meanwhile, the remainder of the studies claiming to spatially analyze coral disease data, either rely on visual examination of disease locations (Grober-Dunsmore et al. 2006; Mayor et al. 2006; Selkoe et al. 2009), or use standard, linear statistics to analyze their spatial data (Selig et al. 2010; Maina et al. 2011). There are several problems with using traditional statistical techniques on spatial data. First, the very nature of these techniques treats the data as if all the points are occurring in the exact same location. Second, many of these statistics are based on underlying assumptions that the data has a normal, homogeneous distribution (Maina et al. 2011); which is an inappropriate assumption considering that both corals (as well as most living creatures) and environmental stressors almost always have heterogeneous spatial distributions (Harley et al. 2006; Ruiz-Moreno et al. 2010; Selig et al. 2010).

The correct use and application of both GIS and spatial analytical methods can provide researchers with powerful new tools for understanding the epizootiology of coral diseases, as well as, improving disease prevention and control. **Table 3.3** provides a summary of some of the different types of information that has the potential to greatly enhance our understanding of coral health, but can be attained *only* through the correct use of GIS and spatial analysis.

Table 3.3 Examples of types of information regarding coral epizootiology that can *only* be attained using geospatial analysis.

- whether or not diseases cluster
- distance scales of disease clusters
- whether these clusters are real or just artifacts of high underlying population density
- where clusters are occurring
- intensity and density information
- spatial prevalence information
- areas with high and/or low clustering levels
- the presence of statistically significant clustering areas
- the ability to take spatial patterns and compare them to environmental factors
- the ability to integrate spatial models with mathematical and/or predictive models
- the ability to locate and investigate and protect areas with increased risk

Chapter 4. Datasets and General Methodology

"Geographic assessments of coral diseases are needed to
understand their local and geographic spatial-temporal variability"
– Weil and Croquer (2009)

4.1 The Study Design

Different types of spatial analysis methods, as well as different parameter settings within each analysis method, can produce noticeably different results. Consequently, poor selection or improper use of a given technique can lead to inaccurate representations of the spatial distribution, resulting in false interpretations of the disease. For this reason, I first performed a comprehensive review of many analytical techniques commonly used by spatial epidemiologists.

Following Cai et al. (2011), the performance, accuracy, and effectiveness of each type of analysis was assessed using an artificial dataset with known cluster locations. In order to ensure that the scale and spatial distribution of the artificial data would be similar to that of an actual coral disease dataset, I created the artificial dataset using the geographic and biologic attributes of data from an actual coral disease outbreak that occurred in the US Virgin Islands (USVI). I then used the results from each of the analyses performed on the artificial cluster dataset to develop a geospatial analytical protocol for coral epizootiology. I then used this protocol to spatially analyze the original coral disease dataset.

4.2 Study Site and Datasets

4.2.1 Buck Island (BUIS) Study Site and White-Band Disease (WBD) Coral Dataset

The Buck Island Reef National Monument was used as the study location for all geospatial analyses presented in this dissertation (**Figure 4.1**). In order to preserve the *Acropora palmata* barrier reef surrounding Buck Island (BUIS) the area was designated a National Monument in 1961; forty years later, in 2001, the park boundaries were expanded from the initial 356*ha* to 7,695*ha* (Causey et al. 2002; Mayor et al. 2006).

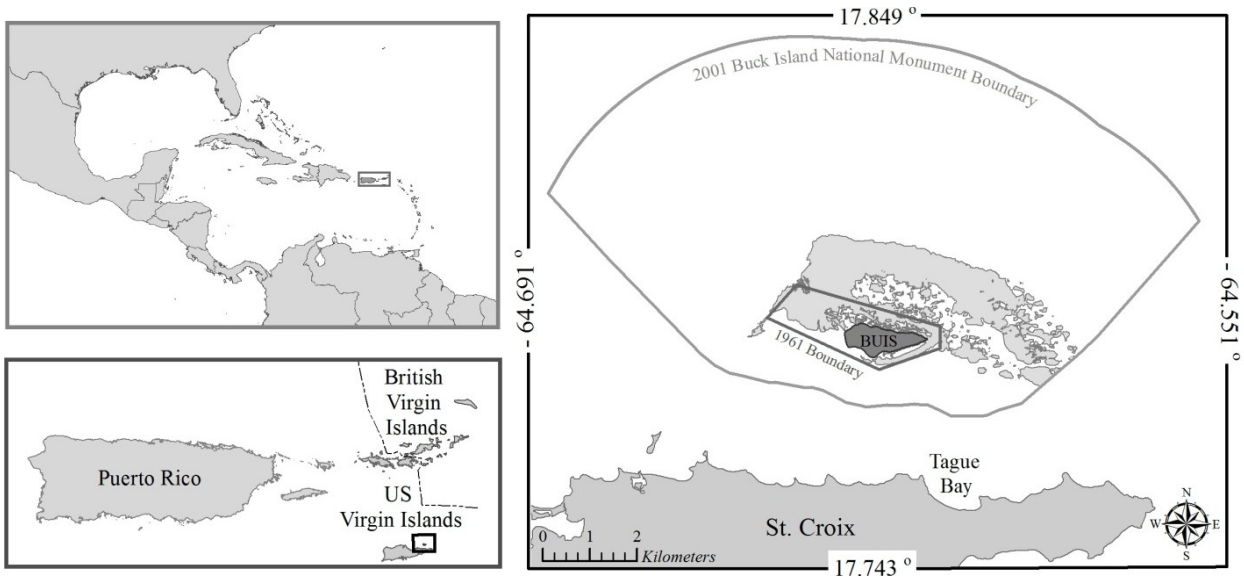


Figure 4.1 The location of the Buck Island (BUIS) Reef National Monument, in relation to the rest of the Caribbean. Note: this figure was adapted from Figure 1 in Lentz et al. 2011.

The coral disease dataset was originally compiled during the summer of 2004 by the US National Park Service for a study by Mayor et al. (2006), which examined the distribution and abundance of *A. palmata* and the prevalence of white-band disease (WBD) around Buck Island (BUIS). In order to increase the chances of locating this already threatened species, Mayor's study only looked at habitats suitable for *A. palmata*, limiting the survey region to hard-bottom areas shallower than 10*m* (depicted as the light grey irregular polygon surrounding BUIS in **Figures 4.1**-**4.3**). Mayor et al. (2006) used *ArcView 3.3* to generate 675 survey points that were randomly distributed throughout the survey region. However, 58 of these points were excluded because they were either located in depths greater than 10*m* or on emerging reefs which could not be surveyed by divers as the tops of the reefs were either at or above the sea surface. Snorkeling teams performed 10 by 25*m* transect surveys at the remaining 617 sites, of which only 375 of the surveyed transects contained *A. palmata* colonies (**Figure 4.2**). These 375 transects contained a total of 2,492 *A. palmata* colonies, 44 of these transects contained 69 colonies with WBD (see **Table 4.1**).

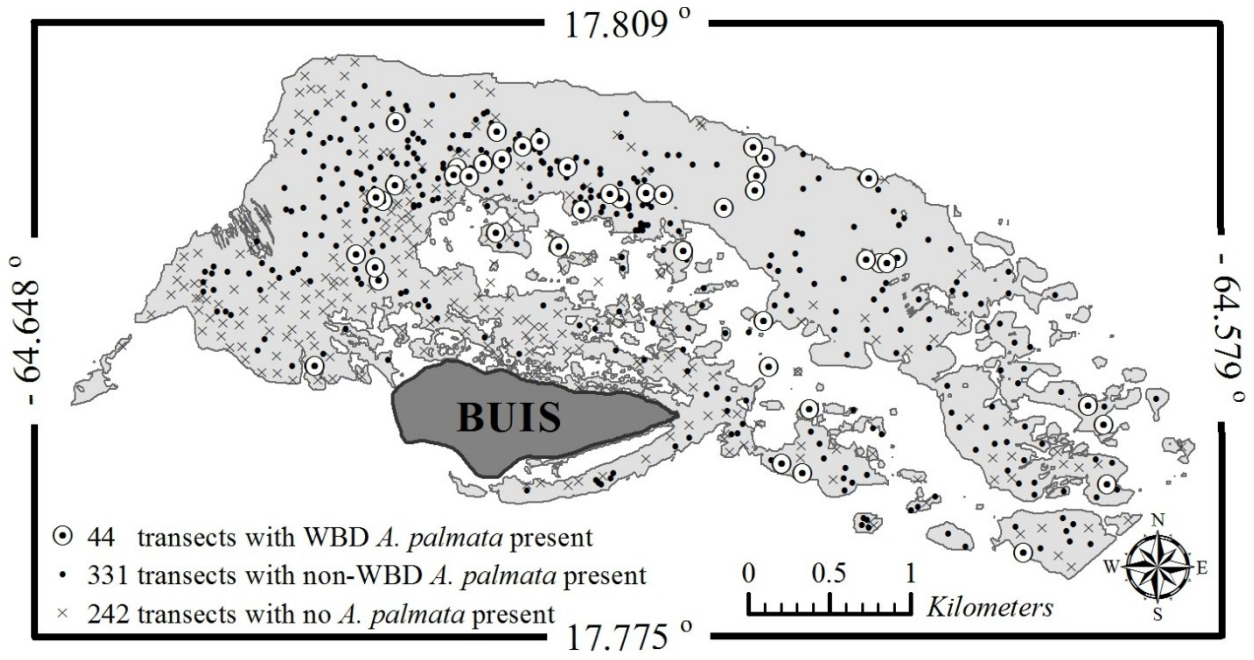


Figure 4.2 The locations of the 617 transects surveyed by Mayor et al. (2006). The light grey region surrounding Buck Island (BUIS) represents hard-bottom substrate < 10*m* deep.

Given that location data had only been recorded for the transects and not for each individual colony, colony-level analyses could only be performed by weighting each transect by the number of colonies within it. In order to see what affect, if any, the colony weighting might have on the resulting spatial patterns, two versions of Mayor's original dataset were created: first, the "colony-level" dataset in which each transect was "weighted" by the number of *A. palmata* colonies within it; and second, the "transect-level" dataset in which all the transects were given the same weight, regardless of the number of colonies within them, a "non-weighted" analysis. Summary statistics for both the transect-level (non-weighted) and colony-level (weighted) versions of the dataset are provided in **Table 4.1**.

4.2.2 Artificial Cluster Dataset

I designed the artificial cluster dataset to have similar geographic properties and biologic attributes to that of Mayor et al.'s (2006) WBD dataset using the following procedure. First, the *Hawth's Tools extension* was used in *ArcMap 9.3.1* to generate four random point locations

Table 4.1 Presence/Absence information can be obtained from the "Transect-Level Data," which shows the total number of transects surveyed with and without white-band disease (WBD) present. The "Colony-Level Data" provides summary statistics for the *A. palmata* colonies found within each of the transects.

Transect-Level Data		**Colony-Level Data**				
Transects containing *A. palmata* colonies	Total # of Transects	Total # of Colonies	# of Colonies per transect			
			Min	Max	Mean	S.D.
with WBD present	44	69	1	6	1.57	1.16
with no WBD present	331	2,423	1	40	6.48	5.87
with & without WBD	375	2,492	1	40	6.65	5.99

within the polygon of surveyed habitat (**Figure 4.3A**), and around which clusters were generated in a circle with a pre-defined radius of 50*m*, 100*m*, 250*m*, and 500*m*. These radii were chosen in order to test the accuracy of the spatial analysis software on detecting clusters of different sizes within the same dataset. The radii were assigned to the cluster centers based on their associated Cluster ID number (**Figure 4.3B**). Since Mayor's study found WBD in 44 of the 375 transects containing *A. palmata*, 11 points were randomly generated within each of the 4 cluster boundaries resulting in a total of 44 clustered points (**Figure 4.3C**). The transects without WBD present were simulated by generating 331 random point locations within the overall study area (**Figure 4.3D**). The end result was an artificial dataset made up of the same number of randomly distributed points as the number of transect locations from Mayor et al.'s (2006) dataset, and within the same geographic area (**Figure 4.3E**).

A "weighted" version of the artificial dataset was created to simulate the colony-level version of Mayor et al.'s (2006) dataset. To do this the colony-level information from Mayor et al.'s transects was blindly assigned to each of the artificial point locations, such that the 44 artificially clustered point locations would have 69 case events within them and the 331 randomly distributed control point locations would have 2,423 control events within them. Thus, the descriptive statistics are the same for both the weighted version of the artificial dataset and the colony-level version of Mayor et al.'s (2006) dataset shown in **Table 4.1**.

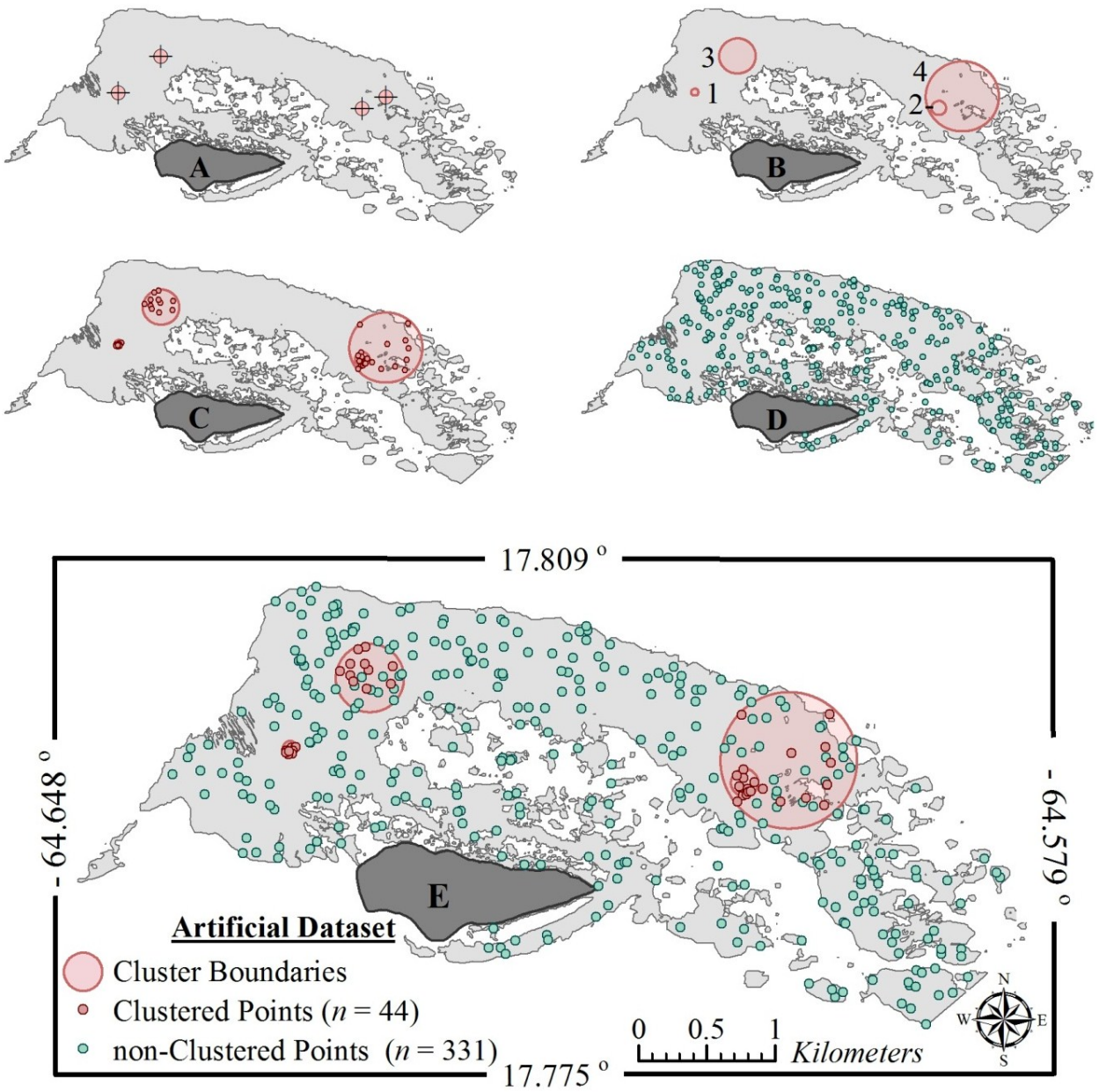


Figure 4.3 Creation of the artificial cluster dataset. (**A**) First, 4 randomly located cluster centers were created within the study area. (**B**) Cluster boundaries were then created by assigning radii to cluster centers based on the ascending order of their size and ID number's respectively, resulting in the following cluster-radii combinations: cluster 1-50*m* radius, 2-100*m*, 3-250*m*, and 4-500*m*. (**C**) Next, 11 points were randomly distributed within each of these cluster boundaries, resulting in a total of 44 clustered points. (**D**) Last, 331 non-clustered points were randomly distributed within the study area. (**E**) The completed artificial cluster dataset.

4.3 Spatial Analysis Software Used

4.3.1 *ArcGIS 9x* and Other ESRI Software[1]

The primary GIS software used throughout this dissertation is the Education Edition of *ArcEditor 9.3.1*. *ArcEditor* is one of many products in the *ArcGIS Desktop* software product line developed by Scott Morehouse with the Environmental Systems Research Institute (ESRI), Inc. (Ormsby et al. 2001; Longley et al. 2005; Lo and Yeung 2007). Within *ArcEditor*, the *ArcMap* application was used to create, analyze, and display the spatial data used throughout this dissertation (Maher 2010). Some of the analyses performed also required the use of ESRI's *Spatial Analyst* and *Geostatistical Analyst* extensions, which are included in the Education Edition of *ArcEditor*. The student edition of *ArcView 3.3*, which is an older version of the ESRI GIS software, was also used in conjunction with the 3x version of the *Spatial Analyst* extension to estimate a few of the statistics used during the analysis performed in Chapter 5.

ESRI's software, and the majority of the GIS software associated with it, is designed to be used primarily, if not exclusively, on Microsoft Windows-based operating systems. The following are either GIS extensions that can be installed directly into the *ArcMap* (or *ArcView 3x*) interface, or are stand-alone programs that produce output that can be imported into a GIS environment to be projected spatially.

4.3.2 *CrimeStat*[2]

CrimeStat is a standalone program that was developed to analyze the locations of crime incidents (Levine 2007; Levine and Associates 2009). The National Institute of Justice funded the development of this spatial statistics program by Ned Levine and his associates (Curtis and

[1] ESRI software is available for purchase online at: *http://esri.com/products/*.
ESRI also offers discounts for students and educators; as well as, free 60-day trial evaluations of the most recent versions of *ArcEditor* and many of the extensions for *ArcGIS Desktop*. Note all of the tools and methods described in this dissertation are also available in the newer version of *ArcGIS* (*ArcGIS 10x*).

[2] *CrimeStat* is publically available as a free download at: *http://www.icpsr.umich.edu/NACJD/crimestat.html*

Leitner 2006; Levine and Associates 2009). *CrimeStat* is not restricted to the type of incidence data for which it was created, but rather can be used for any numerator data regardless of whether or not there is denominator data included in the dataset. **Table 4.2** provides a comprehensive list of the types of analyses offered within the *CrimeStat* program.

Table 4.2 Types of analysis available in *CrimeStat® III*

Distance Measurements
- Direct distance
- Indirect distance
- Network distance

Spatial distribution
- * Mean Center
- * Standard distance deviation
- * Standard deviational ellipse
- * Median center
- * Center of minimum distance
- * Directional mean and variance
- * Convex Hull
- Moran's I spatial autocorrelation index
- Geary's C spatial autocorrelation index
- Moran Correlogram

Distance Analysis
- * Nearest neighbor analysis
- * Ripley's K statistic
- Assign primary points to secondary points
- Within primary file distance matrix
- Between primary & secondary file distance matrix
- Between primary file & grid distance matrix
- Between secondary file & grid distance matrix

Hot spot Analysis
- Mode
- Fuzzy mode
- Nearest neighbor hierarchical clustering
- Risk-adjusted nearest neighbor hierarchical clustering
- Spatial & temporal analysis of crime routine (STAC)
- K-mean clustering
- Anselin's local Moran test

Interpolation
- * Single variable kernel density interpolation
- * Dual variable kernel density interpolation

Space-time Analysis
- Knox index
- Mantel index
- Correlated walk model

Journey-to-Crime Analysis
- Calibrate Journey-to-crime function
- Journey-to-crime estimation
- Draw crime trips

Crime Travel Demand: Trip Generation
- Skewness diagnostics
- Calibrate model
- Make prediction
- Balance predicted origins & destinations

Crime Travel Demand: Trip Distribution
- Calculate observed origin-destination trips
- Calibrate impedance function
- Calibrate origin-destination model
- Apply predicted origin-destination model
- Compare observed & predicted origin-destination trips

Crime Travel Demand: Mode Split
- Calculate mode split

Crime Travel Demand: Network Assignment
- Check for one-way streets
- Create a transit network from primary file
- Network assignment

*indicates the specific functions, tools, and types of analyses that are used in this dissertation

Note: this table was adapted from lists on pages 1.2 – 1.4 of Levine and Associates (2004) *CrimeStat III* manual.

4.3.3 *Disease Mapping and Analysis Program* (*DMAP*)[3]

The *Disease Mapping and Analysis Program* (*DMAP*) is a stand-alone program that was created by the University of Iowa's Department of Geography (Rushton and Lolonis 1996; Rushton et al. 1996). *DMAP* was initially developed to study infant mortality and identify possible clustering of infant deaths (Rushton and Lolonis 1996; Rushton et al. 1996). It has since been used in several other types of health and local area investigations (Rushton et al. 2004; Curtis and Lee 2010; Curtis et al. 2010; Curtis and Mills 2011; Lentz et al. 2011). This program is designed to smooth the disease-rate surfaces and then identify significant rates of disease clustering using Monte Carlo simulations (Rushton and Lolonis 1996; Curtis and Leitner 2006).

Typical data input requires numerator and denominator location data, in which the numerator is the incident or event of interest (i.e. diseased corals) and the denominator is the underlying population in which the incident has occurred (Cai et al. 2011). *DMAP* first aggregates all of the point level data to a circle or "filter" centered on a grid intersection point, with the grid covering the entire study area (**Figure 4.4**). The numerator (case) and denominator (population) points are combined to create a "rate" for each filter. Here, as in many other types of spatial analysis, the term "rate" is not defined using as the commonly accepted definition of "amount of change over time." Instead, the term is used to describe the frequency of one thing relative to another within a given time period (Meade and Earickson 2005), and is calculated as $Rate = \frac{number\ of\ Case\ events\ in\ a\ given\ population\ for\ a\ specified\ place\ and\ time}{Total\ underlying\ population\ at\ risk\ during\ the\ specified\ place\ and\ time}$, in which the "specified place" refers to the space inside a given spatial filter (see **Figure 4.4D**).

It is important to note that spatial filters should be large enough to cover multiple-grid intersections, allowing for points to be included in multiple rate calculations, and thus smoothing the rate surface which eliminates hard (and often artificially defined) aggregation breaks.

[3] *DMAP* is publically available as a free download at: *http://www.uiowa.edu/~geog/health/index11.html*

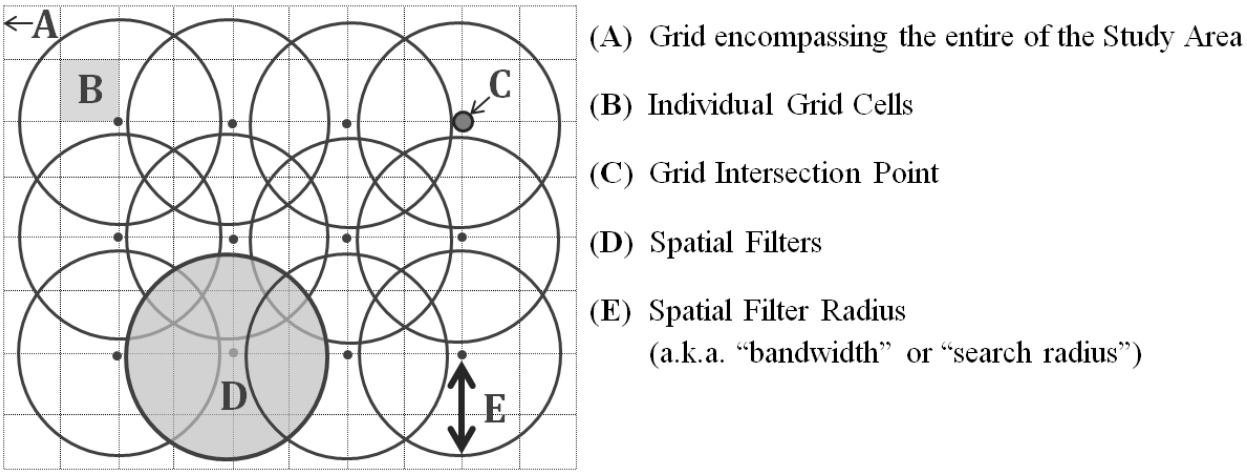


Figure 4.4 Illustration depicting how *DMAP* applies spatial filters to both the numerator (case) and denominator (population) data. Note: this figure was adapted from Figure 3 on page 721 of Rushton and Lolonis (1996).

Once the rates of disease clustering have been found, a Monte Carlo simulation can be done in *DMAP* to identify any areas with significant rates of disease clustering. The Monte Carlo simulation is based on the actual locations of the real data, with a probability added for each healthy individual becoming diseased (this probability was based on the total study area disease rate). A Monte Carlo simulation re-creates this disease surface "*n*" times (for this study *n*=1,000), creating a simulated distribution against which the actual disease surface is compared. If, for example, the disease rate in one filter is actually higher in 990 out of the 1,000 simulation runs, one can be 99% confident (equivalent to a *p*-value of 0.01) that the revealed rate, or hotspot, did not occur by chance alone.

4.3.4 *OpenGeoDa*[4]

OpenGeoDa is a standalone program that was developed by Luc Anselin to perform exploratory spatial data analysis (ESDA) on lattice data (Anselin 2003; Maguire et al. 2005; Anselin et al. 2006a; Leitner and Brecht 2007). The program was designed to provide it's users

[4] *OpenGeoDa* is publically available as a free download at: http://geodacenter.asu.edu/software

with an intuitive and interactive path through the following aspects of ESDA: (1) simple mapping and geovisualization; (2) spatial data manipulation and transformation; (3) spatial autocorrelation; and (4) spatial regression (Anselin 2003; Maguire et al. 2005; Anselin et al. 2006a; Leitner and Brecht 2007). **Table 4.3** provides a more comprehensive list of the analytical categories, as well as, the functions associated with each category (Anselin et al. 2006a). Unlike many of the other GIS programs currently available, *OpenGeoDa* was designed to provide its users with a uniquely interactive environment that combines maps with statistical graphics through its use of dynamically linked windows (Anselin 2003; Maguire et al. 2005; Anselin et al. 2006a; Leitner and Brecht 2007).

It is also important to note that the analyses provided in *OpenGeoDa* are intended to be used with lattice data (Anselin 2003), which is when observations are represented as continuous polygonal surface made up of discrete spatial objects. Discrete data refers to areas in which there is no uncertainty regarding their location (such as the polygonal representation of states or counties), as opposed to point-based event or sample locations which generally have less certainty as to their exact location (Maguire et al. 2005)

4.3.5 *Hawth's Tools* Extension[5,6]

Hawth's Tools is a free extension for ESRI's *ArcMap* (Beyer 2004). The extension was developed by Hawthorne Beyer, and designed to provide users (primarily ecologists) with numerous types of spatial analyses and functions that tend to be more difficult to perform using the functions provided by default in *ArcMap* (Beyer 2004). **Table 4.4** provides a comprehensive list of the tools and functions provided by the extension.

[5] [5] *Hawth's Tools* is a free *ArcMap* extension & is available for download at: *http://www.spatialecology.com/htools/*

[6] [6] During the writing of this dissertation *Hawth's Tools* became a legacy product. Given that the extension is no longer being updated and will soon be formally discontinued, it is not supported in the newest version of *ArcGIS 10x*. This extension has been replaced by a new software package called the "Geospatial Modeling Environment" (GME), which offers all the same tools but in a more flexible environment. GME is a free stand-alone program & is available for download at: *http://www.spatialecology.com/GME/*

Table 4.3 Analytical Categories and Functions available in *OpenGeoDa*

Category	Functions
Spatial Data	* data input from shape file (point, polygon) data input from text (to point or polygon shape) * data output to text (data or shape file) create grid polygon shape file from text input centroid computation * Thiessen polygons
Data Transformation	variable transformation (log, exp, etc.) queries, dummy variables (regime variables) variable algebra (addition, multiplication, etc.) spatial lag variable construction rate calculation and rate smoothing data table join
Mapping	generic quantile choropleth map standard deviational map percentile map outlier map (box map) circular cartogram map movie conditional maps smoothed rate map (EB, spatial smoother) excess rate map (standardized mortality rate, SMR)
Exploratory Data Analysis (EDA)	histogram * box plot * scatter plot parallel coordinate plot three-dimensional scatter plot conditional plot (histogram, box plot, scatter plot)
Spatial Autocorrelation	* spatial weights creation (rook, queen, distance, k-nearest) higher order spatial weights spatial weights characteristics (connectedness histogram) Moran scatter plot with inference bivariate Moran scatter plot with inference Moran scatter plot for rates (EB standardization) * Local Moran significance map * Local Moran cluster map * bivariate Local Moran Local Moran for rates (EB standardization)
Spatial Regression	OLS with diagnostics (e.g., LM test, Moran's I) Maximum likelihood spatial lag model Maximum likelihood spatial error model predicted value map residual map

*indicates the specific functions, tools, and types of analyses that are used in this dissertation

Note: the above table is adapted from Table 1 on page 9 of Anselin et al. (2006a)

Table 4.4 Analytical tools and functions provided by the *Hawth's Tools* extension

Analysis Tools
- Intersect Point Tool
- Distance Between Points (within Layer)
- Distance Between Points (between Layers)
- Count Points In Polygons
- Polygon In Polygon Analysis
- Sum Line Lengths in Polygons
- Line Raster Intersection Statistics
- Enumerate Intersecting Features
- Line Metrics

Sampling Tools
- Create Random Selection
- Random Selection Within Subsets
- *Generate Random Points
- Generate Regular Points
- Conditional Point Sampling Tool
- *Create Vector Grid (lines/polygons)
- Create Sample Shapes (various shapes)
- Generate Random 3D Points

Animal Movements
- Create Minimum Convex Polygons
- Calculate Movement Parameters
- Convert Locations To Paths
- Convert Paths to Points
- CRW Simulation I
- CRW Simulation II

Kernel Tools
- Fixed Kernel Density Estimator
- Batch Fixed Kernel Density Estimator
- Percent Volume Contour

Other Tools
- Digitize XY Coordinates
- Set/Zoom To View Extent Bookmark

Raster Tools
- Clip Raster
- Clip Raster By Polygons
- Landscape Characterization (fast)
- Extract Raster Edge
- Thematic Raster Summary (by polygon)
- Zonal Statistics ++ (by polygon)
- Spatial Replace Tool
- Maximum Grid Separation Tool
- Cellular Automata (1D x Time)
- Grid Spread (Cellular Automata)
- Raster Pixel Type Conversion

Table Tools
- Add Area Field To Table
- Add Length Field To Table
- Add XY To Table
- List Unique Values
- Sum Values
- Delete Multiple Fields
- Add XY Line Data (creates line layer)
- CSV Management Tool

Vector Editing Tools
- Create Buffers (Retain Attributes)
- Vector Rotation and Shifting Tool
- Snap Points To Lines Tool
- Intersect Lines (Make Points)
- Split Vector Layer By Unique Value Field

Specialist Tools
- River Sample Extraction
- Point Redistribution Tool
- PLSS Point Finder
- Julian Day Lookup

*indicates the specific functions, tools, and types of analyses that are used in this dissertation

Note: this table is adapted from the list provided by the developer of the extension, Hawthorne Bayer; available online at: *http://www.spatialecology.com/htools/tooldesc.php*

4.3.6 *Home Range Extension* (*HRE*)[7]

The *Home Range Extension* (*HRE*) is a free extension for ESRI's *ArcView 3x* (Rodgers and Carr 1998). The extension was developed by Arthur Rodgers and Angus Carr and funded by the Center for Northern Forest Ecosystem Research at the Ontario Ministry of Natural Resources (Rodgers and Carr 2002). *HRE* was designed to enable and facilitate the study of the home ranges of animals in *ArcView* through the use of minimum convex polygons (MCPs) and different types of kernel analyses (Rodgers and Carr 2002).

4.3.7 *SaTScan*[8]

SaTScan is a stand-alone program that was developed by Martin Kulldorff to analyze spatial, temporal, and spatio-temporal data of health events using scan statistics (Curtis and Leitner 2006; Kulldorff 2006). The program was designed to: (1) map diseases; (2) detect possible spatial, temporal, or spatio-temporal clustering of a disease and determine whether or not the clustering is statistically significant; (3) determine if a disease is randomly distributed in space or time; (4) predict the locations of future disease outbreaks through the use of prospective analyses (Kulldorff 2006; Kulldorff 2010).

SaTScan uses a number of models depending on the type of data being analyzed (**Table 4.5**). A grid file encompassing the overall study area may be specified, as was the case in *DMAP*; or if no grid is specified the locations of the denominator data (coral locations) will be used to define the centroids of the spatial scan statistics (Kulldorff 2006). Next a circular window will be used by the spatial scan statistic to search the entire study area, centering on each grid point with a radius extending outwards (Wang 2006). The radius of the scanning window is

[7] *Home Range Extension* (*HRE*) is a free *ArcView 3x* extension that is publically available for download at: *http://flash.lakeheadu.ca/~arodgers/hre/* or *http://blue.lakeheadu.ca/hre/*

[8] *SaTScan* is publically available as a free download at: *http://www.satscan.org*

Table 4.5 The eight different Probability Models provided in *SaTScan*

Probability Model	Description
**** Poisson Model*** (Discrete-version)	The *discrete Poisson model* should be used when the background population reflects a certain risk mass such as total person years lived in an area. The cases are then included as part of the population count.
****Bernoulli Model***	The *Bernoulli model* should be used when the data set contains individuals who may or may not have a disease and for other 0/1 type variables. Those who have the disease are cases and should be listed in the case file. Those without the disease are 'controls', listed in the control file. The controls could be a random set of controls from the population, or better, the total population except for the cases. The *Bernoulli model* is a special case of the *ordinal model* when there are only two categories.
Space-Time Permutation Model	The *space-time permutation model* should be used when only case data are available, and when one wants to adjust for purely spatial and purely temporal clusters.
Multinominal Model	The *multinomial model* is used when individuals belong to one of three or more categories, and when there is no ordinal relationship between those. When there are only two categories, the *Bernoulli model* should be used instead.
Ordinal Model	The *ordinal model* is used when individuals belong to one of three or more categories, and when there is an ordinal relationship between those categories such as small, medium and large. When there are only two categories, the *Bernoulli model* should be used instead.
Exponential Model	The *exponential model* is used for survival time data, to search for spatial and/or temporal clusters of exceptionally short or long survival. The survival time is a positive continuous variable. Censored survival times are allowed for some but not all individuals.
Normal Model	The *normal model* is used for continuous data. Observations may be either positive or negative.
Continuous Poisson Model	The *continuous Poisson model* should be used when the null hypothesis is that observations are distributed randomly with constant intensity according to a homogeneous Poisson process over a user defined study area.

* Indicates the specific probability models that are used in this dissertation.

Note: the above information provided in the Description column was taken directly from page 45 of Kulldorff's most recent *SaTScan* User's Guide (Kulldorff 2010).

continuously changing from 0 to the upper limit specified by the user (Kulldorff 2006; Wang 2006). The analysis will ultimately create an infinite number of circles, with each circle representing a possible coral disease cluster. These clusters are determined by the spatial scan statistics estimation of whether there is a statistically significant risk of coral disease inside each circle.

4.3.8 *XTools Pro* 8.0 Extension[9]

*XTool*s Pro is a free extension for ESRI's ArcMap. XTools Pro is a shareware extension designed by Data East, LLC. The extension is designed to provide its users with a comprehensive set of tools for the spatial analysis of vector-based data, shape conversions, table management; and various geoprocessing tools (see **Table 4.6**).

Table 4.6 Analytical tools and functions provided by the *XTools Pro 8.0* extension

Data Management Tools
- Create Feature Class/Table
- Create File OpenGeoDatabase
- Create Personal OpenGeoDatabase
- Delete Dataset
- Change Datasources

Feature Conversions
- Transfer/Convert Features
- Convert Multipart Shapes to Single Parts
- Convert Polygons to Polylines
- *Convert Features to Points
- Make Polygons from Polyline
- Make Polygons from Points
- Make Polylines from Points
- Convert Graphics to Shapes
- Shapes to Centroids
- Split Polylines or Polygons
- Smooth Polylines
- Split Layer by Attributes

Layer Operations
- Create Intersection Points
- Erase Features
- Identity
- Update Polygon Layer

Table Operations
- *Calculate Area, Perimeter, Length, Acres and Hectares
- Add XYZ Coordinates
- Aggregate Features/Records
- Table Restructure
- Export Data to MS Excel
- MultiDelete Fields
- Table Statistics
- Export Table to Text or to HTML
- Find duplicates
- Sort Features/Records

Surface Tools
- Convert Grid to Contour
- Extract Raster Values
- Create Grid from Contours
- rom Points
- en Polygons

Go To tools
- Go to Google Maps
- Go to Google Earth
- Go to Microsoft Bing Maps
- Go to ArcGIS Explorer

Miscellaneous General Usage Tools
- Open Attribute Table
- Smart Add Data
- Show Nodes
- Convex Hull
- Extract Map
- Create Fishnet
- Identify Pro
- Feature Report
- Start Editing Selected Layer
- Show Directions
- Export Data to KML
- Auto Save MXD
- Callout Identify
- Import Data from KML
- Catalog tab & Catalog dockable window in ArcMap
- Edit Metadata
- View Metadata
- Synchronize Metadata
- MXD Info
- Multiple Map Layouts
- Copy Layer Properties

*indicates the specific functions, tools, and types of analyses that are used in this dissertation

Note: this table is adapted from the list provided on the *X Tools Pro* website: *http://www.xtoolspro.com/tools.asp#crt*

[9] *XTools Pro* is a free *ArcMap* extension that is publically available for download at: *http://www.xtoolspro.com/*

Chapter 5. Performing Exploratory Spatial Data Analysis on the Artificial Dataset

"Because so many marine invertebrates, especially corals, are sessile, the use of spatial distributions of diseased and dead animals has the potential to become an important tool in uncovering mechanisms of transmission, disease spread, and the role of host resistance in patterns of mortality from disease."
– Jolles et al. (2002)

5.1 Introduction to Exploratory Spatial Data Analysis (ESDA)

A typical spatial epidemiological investigation would employ an exploratory analysis whereby patterns are identified. This allows for potential associations to be identified and for more traditional hypotheses to be tested. Exploratory spatial data analysis (ESDA) is an important component of GIS based investigations of disease. This is in part due to the large amounts of spatial data that may be important in identifying disease presence, and partly because of the unknown spatial characteristics of many disease systems (Anselin 1995; Fotheringham 1998; Anselin 2003; Curtis et al. 2010; Ratcliffe 2010). For example, clusters of diseased individuals (hotspots) revealed during the ESDA can also be further examined using more traditional epidemiological methods to investigate the epidemiology, etiology, pathology, and pathognomy of the disease (Berke 2004). This hotspot-to-causation approach enables a more comprehensive and less subjective way for epidemiologists to examine the spatial aspects of both the disease transmission and infection rates (Chaput et al. 2002).

The six most common categories of ESDA are as follows: (1) Mapping and Visualizing the data; (2) Point Pattern Analysis; (3) Spatial Filtering and Smoothing; (4) Spatial Scan statistics; (5) Spatial Autocorrelation; and last (6) Spatial Regression. While it is usually a good idea to perform more than one type of analysis on a given dataset, it is often not practical, or appropriate, to perform numerous types of analysis from each of the above categories. For this

reason, one of the objectives of this dissertation is to develop several analytical protocols, which can be used by researchers throughout the world to study coral health. Specifically these protocols will be designed for researchers with little to no background in GIS and/or spatial analysis. Ultimately, my goal is to provide researchers with the necessary tools and information needed to perform the most accurate and powerful types of geospatial analysis possible based on the data they have available.

Developing the different protocols required not only a thorough background on each of the ESDA categories, but also the individual types of analysis within each category. I also needed to understand how the choice of different program settings and/or spatial parameters would influence the accuracy of the resulting analysis. However, without a priori knowledge of the spatial nature of the given coral disease, I would be unable to determine the accuracy of the different analyses and a limited ability to assess their strengths and weakness. For this reason, all of the initial exploratory analyses were performed on an artificial dataset with known cluster locations. The use of an artificial dataset should enable not only better assessment of the accuracy of the various cluster detection techniques used by each of the different types of geospatial analysis; but also, study how spatial scale influenced the results of various types of analysis. Additionally, given the design of the artificial dataset (see the previous chapter for a detailed explanation), I should be able to use the artificial dataset to calibrate the different types of analysis to perform optimally on Mayor et al.'s (2006) coral disease data.

The following sections provide detailed explanations of the six ESDA categories. Each section will begin with a general description of the category, including the purpose, general utility, and potential applications for coral disease studies. The types of spatial analysis generally associated with this category are then listed and explained in varying levels of detail – with more detailed information provided for the types of analysis that were used in this

dissertation. Analyses that were not performed in this dissertation, but have the potential to be important in other coral disease studies, are also described here, but in much less detail. Each section will also contain the specific methods used to perform the given analyses on the artificial dataset.

5.2 Mapping and Visualizing Data Methods

5.2.1 Introduction, Purpose, and Importance of the Techniques in This Category

The first step in many types of exploratory analysis often involves various types of visualization techniques because they facilitate the detection of underlying patterns and trends in the data. Bailey and Gatrell (1995) refer to this "ability to be able to 'see' the data being analyzed" as an "essential requirement" in any analysis. They go on to describe these graphical methods of displaying data as the "fundamental tools of the analyst concerned with seeking patterns in data, generating hypotheses and assessing the fit of the proposed models, or the validity of predictions derived from them" (Bailey and Gatrell 1995; page 22).

When dealing with spatial data, maps are one of the most common and powerful graphical visualization methods available. Maps are defined as "any graphical representation of geographic or spatial information" (Wade and Sommer 2006). Using this definition, maps can be considered "the spatial analysts' equivalent to the invaluable scatter plot in non-spatial analysis" (Bailey and Gatrell 1995; page 22). Maps allow the spatial distribution of phenomena (such as the locations of disease corals) to be visually observed (Câmara et al. 2008). In addition, showing *where* an event is occurring may also provide insight as to *why* the event is occurring (Waller and Gotway 2004).

The concept of medical geography developed out of a long history of scientists using maps as tools for detecting potential "causes" of various diseases (Waller and Gotway 2004). A popular historical example is Dr. John Snow's legendary map of Cholera deaths clustered around

a water pump in the mid-1800s (Waller and Gotway 2004; Câmara et al. 2008). A less well known example is Dr. Theobald Palm's use of maps to link the geographic distribution of rickets (a disease caused by vitamin D deficiency) to urban areas with cold wet climates in the late 1800s, (Hardy 2003; Waller and Gotway 2004). More recently, in the mid 1900's, Dr. Harold F. Blum studied the geographic distribution of disease cases leading him to deduce sunlight as a casual factor of skin cancer (Waller and Gotway 2004).

5.2.2 Map Types

Just as there are numerous graphical techniques for displaying non-spatial data (graphs, box plots, scatter plots, etc.), there are also a number of types of mapping and visualization options for displaying spatial data (Bailey and Gatrell 1995; Waller and Gotway 2004). Data of observed locations can be represented as points, in which the x,y coordinates mark the location of the event; or as polygons, in which the x,y coordinates represent the centroid of the surrounding polygon (Anselin 2003). Visualization options for both the map and the representation of the data locations within the map should be based on the nature of the data, the type of study, and the overall purpose of the map (i.e. what message is the map intended to convey). In many cases, it may be best to try several options and compare the results (Waller and Gotway 2004). Additionally, most of the various types of spatial analysis techniques require the input data be in either point or polygon form. For example, the types of analysis used in the Point Pattern Analysis, Spatial Filtering and Smoothing, and Spatial Scan Statistic ESDA categories usually require the input data to be in point form. Whereas, the analyses in the Spatial Autocorrelation and Spatial Regression categories usually require the input data be a in the form of a continuous polygonal surface (more commonly referred to as "lattice data" or "discrete spatial objects").

One of the most common ways of transforming point data into a continuous polygonal surface is to create Thiessen Polygons (Bailey and Gatrell 1995; Anselin 2003; Wade and

Sommer 2006). These polygons are created by taking point data and constructing polygons around each point such that each point now has a polygon associated with it (see **Figure 5.1**). *OpenGeoDa*, the *XTools Pro* extension, and some versions of *ArcMap*, have tools that can be used to transform points to polygons by generating Thiessen Polygons, as well as, converting polygons to points in which each point represents the center of the associated polygon.

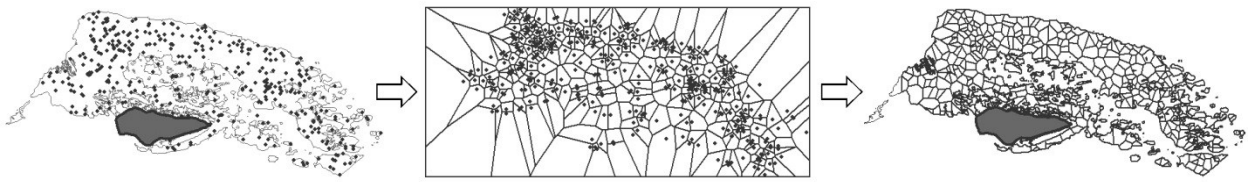


Figure 5.1 Transforming point data into a continuous polygonal surface using Thiessen Polygons. The point locations (left image) are used to construct Thiessen Polygons such that each point is encompassed by a polygon (center image). The polygonal surface is then "clipped" to the boundaries of the study area (image on the right).

5.2.3 Mapping and Visualization Techniques Performed on the Artificial Dataset

The two most common methods of displaying spatial data in a GIS were compared by displaying the artificial dataset as point data in one map and a continuous polygonal surface in the other. In both display methods, the case (artificially clustered) and control (artificially non-clustered) locations were differentiated using the colors red and green respectively. Solid colors were used to depict the non-weighted (case and control locations) versions of the artificial dataset; while gradients of the same two colors were used to depict the weighted versions, with darkening shades indicating the presence of an increased number of individual case or control events at a given location. Additionally, the size of the point symbols used to depict the weighted event locations were also scaled to reflect the number of individuals present.

5.3 Point Pattern Analysis (PPA) Methods

5.3.1 Introduction, Purpose, and Importance of the Techniques in This Category

While the importance of visualizing the spatial distribution of point patterns was recognized early on, as was shown by the historical medical applications mentioned in the previous section, it was not until the 1950s and 1960s that researchers began to seek out methods for statistically analyzing these spatial patterns (Gatrell et al. 1996). Despite the diagnostic breakthroughs that had been made by simply mapping disease occurrences, the first techniques to analyze spatial point patterns were developed not by the medical community, but rather by plant ecologists (Gatrell et al. 1996). During the 1960s and 1970s researchers from other disciplines began adopting some of these methods for describing and studying plant distributions, and applying them to other fields of study, such as urban planning (Gatrell et al. 1996). These early types of spatial analysis were limited by the technology available at the time and in large part had to be computed by hand. Most of the analyses were either centrographic or distance-based statistics that used physical distance measurements to characterize the overall spatial distribution of the points (Haggett et al. 1977; Gatrell et al. 1996).

5.3.2 Common Types of PPA

Most types of point pattern analysis (PPA) are designed for the analysis of just location data, rather than or in addition to attribute data. For this reason PPAs are often performed only on data regarding the "case" events, without taking into account the spatial distribution of the "control" events (where the events of interest are not occurring), or the distribution of the underlying population (both case and control events).

The general objective of PPA is to describe the overall pattern in the spatial distribution of the point locations (**Figure 5.2**). For example, Bo the points appear to have a more clustered, dispersed, or random spatial distribution? Where is this spatial pattern generally occurring?

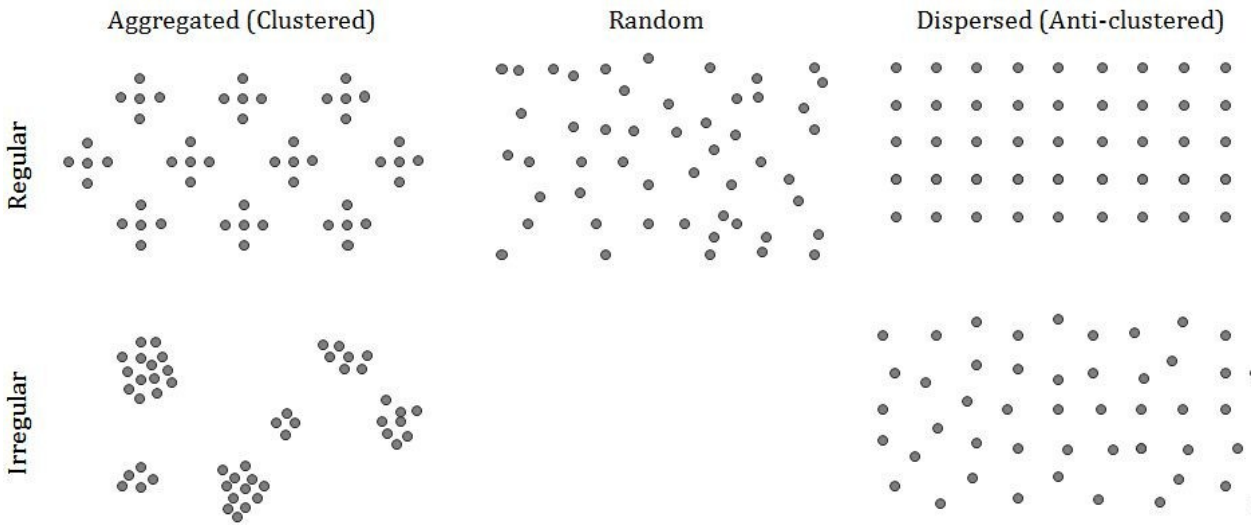


Figure 5.2 Visual examples of common types of spatial distribution patterns for point data. Note: the above figure was adapted from Figure 10.3 on page 381 of Lo & Yeung 2007.

Does the pattern change when the data are analyzed and/or sampled at different geographic scales?

Since many types of spatial analysis deal with point data, they are often included under the PPA category. For example, the following types of analysis deal with point data and are sometimes referred to as "point pattern analyses": different mapping & visualization techniques; centrographic and distance statistics; spatial filtering, smoothing, and interpolation methods (including kernel density estimates); spatial scan statistics; and some spatial autocorrelation methods. For clarity purposes only centrographic and distance statistics will be included under the PPA category; the other analyses will each be described as their own ESDA category.

5.3.2.1 Centrographic PPA Statistics

Centrographic statistics are types of descriptive spatial statistics designed to show the location and distributional area of the overall point patterns (Levine and Associates 2004; Tabangin et al. 2010). While the types of analyses included under this heading do contain locational information, they are often referred to as "global" statistics rather than "local"

statistics (Chainey and Ratcliffe 2005; Tabangin et al. 2010). This is because, centrographic statistics are used to summarize the similarities in the spatial distributions of points rather than the localized differences (see **Table 5.1**).

Table 5.1 The difference between "Global" Spatial Statistics and "Local" Spatial Statistics

Global Spatial Statistics	**Local Spatial Statistics**
Used to emphasize the similarities over space	Used to emphasize the differences over space
Used to search for region-wide trends	Used to search for local exceptions or hotspots
Spatial distribution is assumed to be homogeneous	Spatial distribution is assumed to be heterogeneous
Results are often non-spatial (not-mappable)	Results contain spatial output (mappable)
Results are usually single-value statistics	Results are usually multi-value statistics

Note: the above table has been adapted from Table 5.1 on page 94 of Fotheringham et al. (2000), and Table 1.1 on page 6 of Fotheringham et al. (2002).

The following are common types of centrographic spatial statistics:

Minimum Convex Polygons (*MCPs*), also referred to as "Convex Hull" polygons, are polygons which represent the external boundary of a point distribution. MCPs are created by connecting the outermost points in a spatial pattern, thus the completed polygon serves as the border or perimeter of the points. MCP estimates are quite common in animal movement studies, because they are the oldest and simplest method for calculating the home range of a given animal (Plummer 2003; Lentz 2005).

Mean and median center statistics are used to find the central focal point within a given spatial distribution. Mean center statistics (such as the *Harmonic Mean, Geometric Mean,* and the *Mean Center*) are generally calculated as the means of the *x* and *y* coordinates (Levine and Associates 2004; Smith and Bruce 2008; Tabangin et al. 2010). In a similar fashion, the *Median Center* is calculated as the median of the *x* and *y* coordinates, which can be of use when spatial outliers are influencing the *Mean Center*. *Mean Center of Minimum Distance* is the "point at which the sum of the distance to all other points is the smallest" (Smith and Bruce 2008). In

spatial crime analysis, the *mean center of minimum distance* statistic is often the "single best predictor" of where a serial offender lives, and consequently tends to be more useful than either the *mean* or *median centers* (Smith and Bruce 2008). Standard distance and deviation statistics (such as the *Standard Distance Deviation*, *Standard Deviation of the X & Y Coordinates*, and *Standard Deviational Ellipse* statistics) are often used in conjunction with the mean and median centers to estimate the general distribution of the data around the central focal point (Gatrell et al. 1996).

5.3.2.2 Distance PPA Statistics

Distance statistics are generally used to test hypotheses regarding the spatial distribution of points. The following are the two most common types of distance PPA statistics: *Nearest Neighbor Analysis (NNA)*; and the *Ripley's K* statistic. Both are used to examine "spatial dependence" (clustering or dispersion of points) and are also considered "global" statistics, but unlike centrographic statistics, neither the *NNA* nor the *Ripley's K* statistic have spatial output.

Nearest Neighbor Analyses (NNA) are used to examine spatial dependence and determine whether the points in a given spatial distribution are more clustered or dispersed than would be expected to occur through chance alone (Smith and Bruce 2008). This technique begins by measuring the distance between each point and the point closest to it (its "nearest neighbor"). The mean of these *nearest neighbor* distances is then calculated. The estimated (or "observed") mean nearest neighbor distance is then compared to the mean distance that would be expected based on a random spatial distribution (Smith and Bruce 2008). The result of this comparison is the *Nearest Neighbor Index* (*NNI*). $NNI < 1$ suggests the points are more aggregated (clustered), while $NNI > 1$ indicates the points are more dispersed than would be expected through chance alone, and an $NNI = 1$ indicates "Complete Spatial Randomness" (CSR) in the observed spatial distribution (Bailey and Gatrell 1995; Smith and Bruce 2008).

The *Ripley's K* statistic is a global PPA and spatial autocorrelation statistic, which is used to examine the extent of spatial dependence across several distances (Lentz et al. 2011). In many types of spatial analysis, the scale or resolution of the data tends to have a strong influence on the appearance of the resulting analyses. The *Ripley's K* statistic can be used to study how the spatial dependence present within a given set of points changes across multiple distances (Bailey and Gatrell 1995; Lentz et al. 2011).

Most studies which use the *Ripley's K* statistic, calculate the statistic be using the following linear transformation of the *K*-function:

$$L(d) = \sqrt{\frac{A\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} k(i,j)}{\pi n(n-1)}}$$

where *n* is the total number of point locations, *k* is the number of events (or individuals) present at each location, *A* is the size of the study area, and *d* is the distance over which the spatial autocorrelation is being tested (Bailey and Gatrell 1995; Lancaster and Downes 2004; Wiegand and Moloney 2004; Marcon and Puech 2009; Lentz et al. 2011). Spatial dependence is visually assessed by plotting the resulting *Observed* and *Expected K* values ($L(d)$ and d, respectively) values against distance values (**Figure 5.3**). The *Expected K* values (d) represent the null distribution of complete spatial randomness (CSR), also referred to as the "Poisson distribution." By plotting both the *Observed* and *Expected K* values on the same graph, the *Expected K* values can be used to test the spatial distribution of the *Observed K*s against the null distribution of CSR. Spatial clustering is indicated by the presence of *Observed K* values above the line of *Expected K* values, while observed values that fall below this line indicate spatial dispersion, and observed values that fall along this line indicate random spatial distribution (**Figure 5.3**). Unlike, the other PPA statistics, the spatial distribution of the underlying population at risk can also be accounted for through the use of the difference function (D).

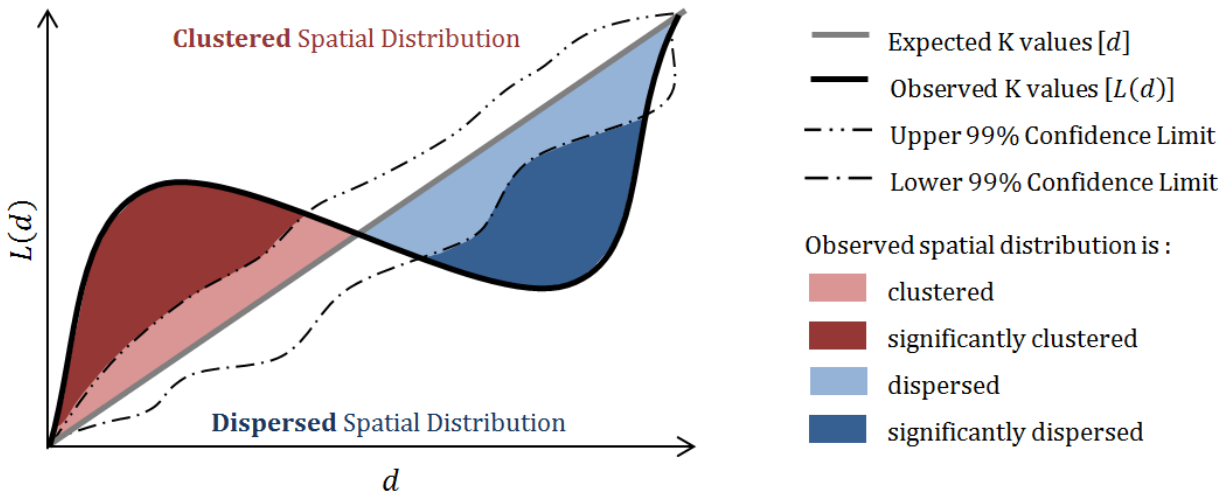


Figure 5.3 How to interpret *Ripley's K* plots

5.3.3 PPA Performed on the Artificial Dataset

CrimeStat III was used to calculate all of the centrographic statistics and to perform the *Nearest Neighbor* distance statistical analyses on the artificially clustered point locations (i.e. non-weighted case locations). These analyses were performed solely on the non-weighted artificial case data, because none of these statistics are designed to incorporate attribute information into their estimates. Thus, given that only the latitude and longitudinal data are being examined, the results would be the same for both the non-weighted and weighted versions of the dataset. Further, since these analyses are not designed to examine or compare more than one spatial distribution, the analyses were only performed on the case data, because what I am most interested in is their ability to correctly identify and detect the artificially defined clusters. The spatial results from the centrographic analyses were brought into *ArcMap* to be displayed and compared visually; while the results of the *NNA* were summarized using tables and graphs.

The *Ripley's K* analyses were all performed using the "Multi-Distance Spatial Cluster Analysis (Ripley's k-function)" tool in *ArcMap*'s spatial statistics toolbox. *Ripley's K* estimates were performed on both the weighted and non-weighted versions of the artificial case and artificial population data. Each of the *Ripley's K* analyses (4 in total) examined distances

ranging from 0 – 1000*m*, in 50*m* intervals, and calculated 99.9% upper and lower confidence intervals by performing 999 permutations. The results of these analyses were then brought into an *Excel* spreadsheet to be analyzed. Normalized and non-normalized ([$L(d)$-d] and [$L(d)$], respectively) *Ripley's K* plots were generated for the weighted and non-weighted versions of the case and population data. Last, the difference function (D) was used to test additional hypotheses regarding the spatial distributions of the case and population data.

5.4 Spatial Filtering and Smoothing Methods

5.4.1 Introduction, Purpose, and Importance of the Techniques in This Category

Since it was first introduced in the early 1950s (Fix and Hodges 1951; Silverman and Jones 1989), Kernel density estimation (KDE) has become one of the "most popular" statistical methods for analyzing both univariate and multivariate data (Danese et al. 2008). KDEs were first developed to "obtain a smooth estimate of a univariate or multivariate probability density from an observed sample of observations" (Bailey and Gatrell 1995). In the early 1980s, the spatial applications of KDEs started to become apparent (Diggle 1983,1985; Danese et al. 2008). Today, KDE are one of the most common types of analysis used to study spatial data. These spatial density estimates are often computed using a process known as "spatial filtering," a type of non-parametric, graphical analysis which calculates the predicted value at a given point based on the values of the surrounding data points (Carlos et al. 2010; Cai et al. 2011). These spatial filters, which are also referred to as "disk smoothers" or "punctual kriging," are used to smooth some of the variability and noise in the dataset without losing the local features of the data, resulting in the creation of smooth, continuous maps of density estimates (Waller and Gotway 2004; Anselin et al. 2006b; Carlos et al. 2010; Cai et al. 2011). One of the most common methods of spatial filtering involves the use of probability distribution functions, known as "kernels" (Danese et al. 2008).

5.4.2 Spatial Kernels and How They Are Used by Spatial Filters to Smooth Data

Spatial filtering is a form of data smoothing in which a three-dimensional kernel probability function moves across the study area filtering and smoothing the underlying point data as it goes. In order to create a continuous spatial density estimate, or density "surface," a fine grid is placed over the entire study area, depending on what type of software is being used the kernel function is either centered on each data point (*CrimeStat*) or centered on each node of the grid (*DMAP*), performing density estimates on the individual point data located in the pre-defined filtering area. This process is described in more detail in the following sections.

5.4.2.1 Spatial Kernels

Kernels are hump-shaped bins placed over each data point (as shown in **Figure 5.4A**), or conversely over each grid point. The area over which the kernel probability function is applied is defined by the size of the bin's radius (h), usually referred to as either "bandwidth" or "filter radius" (Fotheringham et al. 2000,2002). Generally the larger this bandwidth, the more data points are included, and the "smoother" the resulting density map (Williamson et al. 1998,1999; Carlos et al. 2010; Curtis et al. 2010). A visual inspection of these density maps will reveal areas of high disease intensity worthy of subsequent investigation. The shape of the hump (**Figure 5.4B**) indicates how the non-parametric filtering statistic will be applied to the data within the filter radius (Danese et al. 2008). For example, spatial kernels with quartic distributions will put the most weight on the data points closest to the grid point in which the spatial filter is being applied; whereas, spatial kernels with uniform distributions will weight all of the data points within the filter radius equally (Levine and Associates 2004; Smith and Bruce 2008).

5.4.2.2 Single Kernel Density Estimates (KDEs) vs. Dual KDEs

KDEs can be performed as a univariate analysis of a given set of point locations, or as a multivariate analysis of two sets of point locations. The former type of analysis is referred to as

a “single” KDE and is most commonly performed on data regarding the locations of case events, while, the latter is referred to as a “dual” KDE and can be used on various types of “numerator” and “denominator” data. “From an epidemiological perspective, kernel estimation is of most value in estimating the intensity of one type of event relative to another” (Carlos et al. 2010). Such as, comparing the spatial densities of case locations to that of the underlying population at risk; comparing case and control locations; or even case events to a different type of case events using an additive or subtractive dual kernel model (Levine and Associates 2004; Smith and Bruce 2008). When population data are available, dual kernels make it possible to differentiate “real” case clusters and areas (or periods of time) at greater risk (Bithell 1990; Waller and Gotway 2004; Carlos et al. 2010). Conversely, single KDEs of just case data run the risk of identifying false clusters resulting from underlying population dynamics (Bithell 1990; Rushton and Lolonis 1996; Levine and Associates 2004; Waller and Gotway 2004; Anselin et al. 2006b; Smith and Bruce 2008; Carlos et al. 2010).

5.4.2.3 Fixed Distance vs. Adaptive Distance Kernels

As described earlier the area over which the spatial kernel is applied is defined by the size of the filter radius, or, as it is more commonly referred to, “bandwidth.” The size of the bandwidth can be defined in one of two ways: either as a fixed distance (**Figure 5.4C**), or as the number of points to be sampled by each filter (**Figure 5.4D**). There are pro’s and con’s associated with both fixed and adaptive KDEs. When the primary concern of a study is distance, static bandwidths are often more appropriate because the resulting output can be used to define areas of increased risk of exposure based on distances (Carlos et al. 2010). Distance-based kernels are also preferred in situations where a priori information suggests an appropriate distance; for example, studies of vector-based diseases might set the bandwidth to the average distance traveled by the vector.

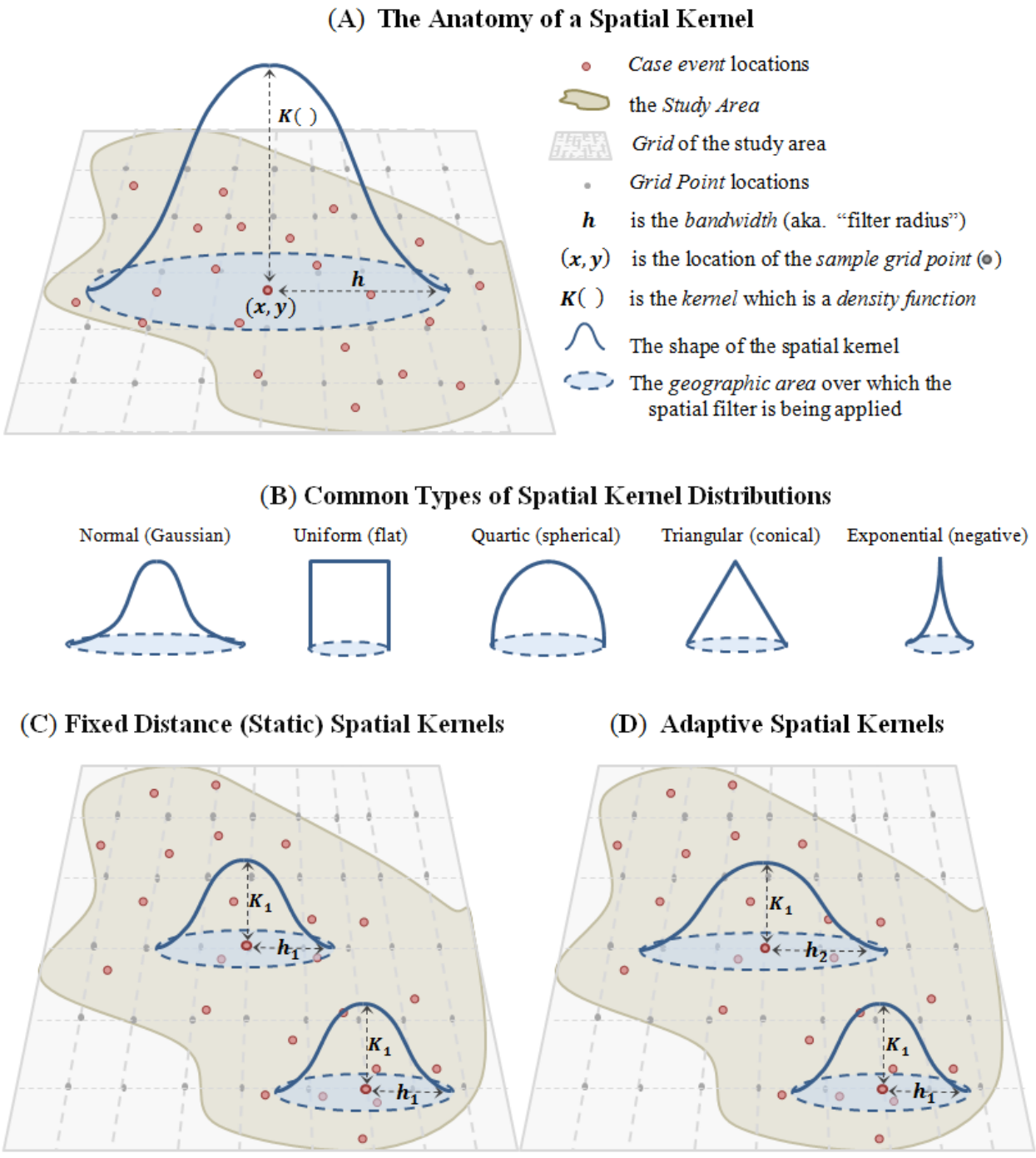


Figure 5.4 Cross-sections of different types of spatial kernels: (**A**) diagram depicting the general anatomy of a kernel-based spatial filter; (**B**) the five most common types of spatial kernel distributions; (**C**) fixed distance (static) bandwidth spatial kernels; and (**D**) adaptive distance spatial kernels. Note the above figure is based on the following sources: Figure 3.4 on page 86 of Bailey and Gatrell (1995); Fotheringham et al. (2002)'s Figures 2.11 and 2.13 on pages 45 and 47 respectively; Figure 3.2 on page 37 of Wang (2006); pages 67-68 of Smith and Bruce (2008); and Figure 4-47 on page 177 of de Smith et al. (2009).

There are two main problems associated with using fixed distance bandwidths: the first is that spatial filters with a fixed size treat high and low density areas the same; and the second, is the user must define the size of the bandwidth (Tiwari and Rushton 2005). Adaptive bandwidths address both problems to varying degrees. Unlike fixed-distance kernels, adaptive KDEs base the size of the bandwidth on the number of point locations rather than geographic distance, and are therefore, less likely to over or under smooth areas with high or low underlying population densities (Gatrell et al. 1996; Talbot et al. 2000; Tiwari and Rushton 2005). Some studies have found that the density rates produced by adaptive kernels have higher statistical stability and provide greater geographic detail when compared to density estimates from fixed-distance kernels of the same data (Tiwari and Rushton 2005; Carlos et al. 2010; Cai et al. 2011). However, adaptive bandwidths are not without fault. For example, density maps based on adaptive kernels will not have the same geographic resolution throughout; thus, a map of disease incidence may show a high rate in a rural area giving the visual impression of local hotspot, when the rate is actually an artifact of the large bandwidth size (Talbot et al. 2000).

The type (fixed vs. adaptive) and size (distance vs. number of point locations) of the bandwidth are very important because different parameter choices often result in dramatically different visual outputs (Gatrell et al. 1996; Tiwari and Rushton 2005; Danese et al. 2008; Carlos et al. 2010; Ratcliffe 2010; Cai et al. 2011). Over the last 20 years, numerous studies have tried to address this issue by creating new ways of calculating the "ideal" fixed or adaptive bandwidth for a given dataset. Unfortunately, there is currently "little guidance available for the novice analyst" on parameter selection (Eck et al. 2005). Some studies recommend researchers experiment with multiple sizes and/or types of bandwidth when using spatial kernels (Wang 2006; Carlos et al. 2010). However, without robust background in both statistical and spatial analysis, this can be quite difficult and time consuming for researchers. Consequently, many

researchers rely on the default setting provided by the software they are using to perform the spatial KDE.

One of the goals of this dissertation is to address this problem by developing recommendations for selecting the most appropriate kernel and bandwidth type based on the data available. To do this I will first provide a review of the most common techniques for estimating the size for both fixed and adaptive bandwidths. I will then use each of these techniques to generate "ideal" bandwidths for the artificial dataset. These bandwidths will then be used to perform single and dual KDEs of the artificial case (numerator) and population (denominator) data. The strengths and weaknesses of each will then be assessed by comparing the results of these analyses to the pre-defined cluster locations.

5.4.3 Spatial Parameter Estimation[1]

Spatial kernels require the user to define the size of two spatial parameters: a grid of the study area with a defined cell size; and the size of the filter radius (also known as "bandwidth"). Some programs also require the user to specify which type of spatial kernel distribution should be used (**Figure 5.4B**), and whether the filter radius should use a fixed distance bandwidth (**Figure 5.4C**) or an adaptive distance bandwidth (**Figure 5.4D**). Spatial parameter selection is a critical first step as the finest bandwidths will lack the potential to identify local areas of disease clustering. Conversely, parameters that are too coarse will overestimate the disease surface and often under estimate the severity of localized clusters (Hall and Marron 1991; Hazelton 1996; Jones et al. 1996; Danese et al. 2008). The size of the grid cell is also important because it is what enables identification of the clustering patterns. Grid cells that are too small cause the interpolation to become jagged, while excessively large grid cells lose the fine-scale detail (Wiegand and Moloney 2004; Chainey and Ratcliffe 2005; Danese et al. 2008; Ratcliffe 2010).

[1] See **Appendix A** for more detailed descriptions on how each of the spatial parameters was calculated

One challenge with coral disease analysis is that the uncertainty surrounding the etiology of a given disease, making it unclear what bandwidth or cell size is needed to adequately capture the disease spreading mechanism (Lentz et al. 2011). In cases where the spatial parameters are not able to be based on a priori knowledge of the epidemiology of the disease being studied, the researcher must generate their own spatial parameters (the grid cell resolution and bandwidth size) based on the dataset itself (Wand and Jones 1995; Danese et al. 2008; Carlos et al. 2010; Lentz et al. 2011). Since there is currently no "universally accepted" way of calculating these data-specific spatial parameters, the most popular techniques associated with fixed and adaptive distance spatial kernels were applied to the artificial dataset so their effectiveness could be assessed and compared.

5.4.3.1 Estimating Spatial Parameters for Fixed Distance Kernels

The following sections describe the methods which were applied to the artificial dataset. The results of each these calculations were then used when the fixed distance spatial filters were applied to the artificial dataset.

Two methods were used to estimate the appropriate grid cell size for the artificial dataset. The first method used an equation proposed by Chainey and Ratcliffe (2005) in which the cell size is "the result of dividing the shorter side of the minimum bounding rectangle (i.e. the shortest of the two extents between maximum *x* and minimum *x*, and the maximum *y* and minimum *y*) by 150." The second method derived the "appropriate" grid cell resolution from the results of the "Visual Calibration Method."

The "Visual Calibration Method" was used to test the capabilities of the spatial analysis techniques based on a controlled dataset producing "known" clusters (Perry et al. 2006; Wiegand et al. 2007; Cai et al. 2011). This method used the artificial dataset to visually assess which spatial parameters most accurately detect the pre-determined artificial clusters. In order to test

the accuracy of spatial filters in identifying these different sized clusters, the *Disease Mapping and Analysis Program* (*DMAP*) was run using 11 different grid cell resolutions ranging from $25m^2$ to $500m^2$, and 11 filter radii ranging from 25*m* to 500*m*, resulting in a total of 121 different analyses. The artificial cluster boundaries were overlaid on each of the *DMAP* analyses in order to visually assess which spatial parameters most closely predict the locations of the artificial clusters. The filter radius from the map that most accurately identified of the locations the artificial clusters will be termed the "Visual Calibration" bandwidth ($\boldsymbol{h_{VC}}$) when used in subsequent analyses.

Five equation-based "Direct bandwidth calculation methods" were employed using equations with various combinations of sample size (n), standard distance (σ), and distance and area measurements based on the study area and data points locations. The first method set the bandwidth equal to the default search radius used by the kernel density tool in *ArcView*'s Spatial Analyst Extension, in which the bandwidth is calculated by dividing the minimum *x,y* by 30 (Williamson et al. 1998,1999). Given that this estimation is performed by the *ArcView* software, the bandwidths based on this method are referred to as "$\boldsymbol{h_{AV}}$" bandwidths. The second method used a technique recommended by Bailey and Gatrell (1995) in which the ideal bandwidth for a given dataset could be attained using the following equation: $0.68n^{-0.2}\sqrt{Study\ Area}$ (Bailey and Gatrell 1995; Williamson et al. 1998,1999); "$\boldsymbol{h_{BG}}$" refers to the bandwidths resulting from this technique. Next, the *Maximal Smoothing Bandwidth* ($\boldsymbol{h_{max}}$) was calculated using the following equation: $h_{max} = 1.144\sigma\ /\sqrt[5]{n}$ (Terrell 1990; Fotheringham et al. 2000). Then the *Optimized Bandwidth* ($\boldsymbol{h_{opt}}$) was calculated using the equation: $h_{opt} = [(2/3n)^{0.25}][\sigma]$ (Terrell 1990; Bowman and Azzelini 1997; Fotheringham et al. 2000; Lentz et al. 2011). Last the *Reference Bandwidth* ($\boldsymbol{h_{ref}}$) was calculated as: $h_{ref} = n^{-1/6}\sqrt{(var_x + var_y)/2}$ (Silverman 1986;

Worton 1989; Carr and Rodgers 2002; Rodgers and Carr 2002; Rodgers and Kie 2010). For each of these 5 methods, the standard distance (σ) was calculated using the Standard Distance Deviation tool in *CrimeStat*, and the sample size (n) was computed as the total number of point locations.

The following "Regression-based Bandwidth Selection Criteria" were also employed: the *Least Squares Cross-Validation* (*LSCV*) criterion; the *Biased Cross-Validation* (*BCV2*) criterion; the *Corrected Akaike's Information Criterion* (*AICc*); the *Least Squares Criterion* (*LSC*); and the *Generalized Cross-Validation* (*GCV*) criterion (Press et al. 1986; Sain et al. 1994; Wand and Jones 1995; Worton 1995; Jones et al. 1996; Carr and Rodgers 2002; Fotheringham et al. 2002; Rodgers and Carr 2002; Rodgers and Kie 2010). The first two bandwidth selection criteria (*LSCV* and *BCV2*) were performed using the *Home Range Extension* (*HRE*) in *ArcView 3.3x* (Rodgers and Carr 1998,2002; Carr and Rodgers 2002). Following Fotheringtham et al. (2002), the "Spatial Statistics toolbox" in *ArcInfo 9.3* was used to perform the regression-based calculations used by the remaining three bandwidth selection criteria (*AICc*, *LSC,* and *GCV*). Due to the design on the software, the point data had to be in the form of a continuous polygonal surface, so the Thiessen polygons created for Mapping and Visualization methods were used as the input data here. The dependent variable (Y) was set as the case locations and the independent, or "explanatory," variable (X) was set as the locations of all 375 points in the artificial dataset (representing the underlying population at risk). Geographically Weighted Regression (GWR) was used to estimate the *AICc*, *GCV*, and *LSC* values, for 20 bandwidths ranging from 50*m* to 1,000*m* in 50*m* intervals and once with "AICc" selected as the bandwidth method. The output from these analyses were used to calculate the criterion values in their respective equations. These criterion values were then plotted against the 20 bandwidths; the lowest point on each graph indicates the "ideal" bandwidth for each selection method.

Last, the Nearest Neighbor Analysis (*NNA*) tool in *CrimeStat III* was used to calculate the mean nearest neighbor distance values for 20 orders of *K*, resulting in a bandwidth ($\boldsymbol{h_{NNA}}$) estimation for each order of *K* (Williamson et al. 1998; Williamson et al. 1999; Levine and Associates 2004; Levine 2007; Levine and Associates 2009).

5.4.3.2 Estimating the Spatial Parameters for Adaptive Distance Kernels

The two grid cell resolutions estimated for the fixed distance kernels were used for the adaptive KDEs. The filter radius for the adaptive KDEs was calculated as no more than 50% of the sample locations, with the single KDEs using the artificial case data (n_{Case}= 44) and the dual KDEs using the artificial population data ($n_{Population}$= 375) as their sample locations, resulting in the following adaptive distance bandwidths $\boldsymbol{h_{Case}} \leq 22$ and $\boldsymbol{h_{Population}} \leq 187$, respectively.[2]

5.4.4 The Spatial Filtering and Smoothing Analytical Process

The following types of spatial kernel density estimates (KDEs) were performed on the artificial dataset using each of the estimated spatial parameters. Each of the following analyses used a rectangular-shaped grid of the study area defined by the following coordinates: 17.816°N, 64.657°W in the northwest and 17.769°N, 64.572°W in the southeast.

Kernel Density Estimates (KDE) were performed in *CrimeStat III*, as opposed to using the kernel tool available in *ArcMap*'s Spatial Analyst extension, because the *CrimeStat* software offered more options for how to perform the KDE and provided extensive documentation about how each of these options worked. The KDE results output was saved as "ArcInfo Grid/ArcView Spatial Analyst 'ASC'" files. The ASC file format was preferred over the

[2] Note, at the time the adaptive KDEs in this study were being performed, I was unable to find any information in either the primary or secondary literature on how to calculate the appropriate number of points for adaptive KDEs. However, I did find that *SaTScan*'s spatial scanning statistic (described in section 5.5) adjusts the size of the scanning window based on a pre-defined number of point locations, which is how adaptive distance KDEs work. I therefore based my adaptive bandwidth calculation on the default used by *SaTScan* (≤ 50% of the population at risk). Given that the *SaTScan* documentation does not explain what their default is based on, this may be an arbitrary or inappropriate calculation method. Therefore, any future adaptive KDEs should use this calculation method with caution, and only after consulting recently published literature to ensure that better estimation method(s) have not been proposed.

"ArcView 'SHP'" format. This is because when 'SHP' is selected as the output format, *CrimeStat* will export the KDE results in the form of a polygon grid saved as a shapefile (SHP). The shapefile then has to be brought into *ArcMap*, converted to a point grid (using the "Convert Features to Points" tool in the *XTools Pro* extension), and then these points need to be interpolated using the "Interpolate to Raster" tool in the Spatial Analyst extension's toolbar. Whereas, when ASC files are brought into *ArcMap* they are already in raster form.

Single KDEs were performed on non-weighted and weighted versions of the artificial case data, using fixed and adaptive bandwidth kernels with Quartic distributions. The "Relative density" option was selected for the output units for the single KDEs of the case data. "Relative density" divides the absolute case densities by the grid cell area ($50m^2$), resulting in output units of "cases per m^2" rather than "cases per grid cell" (which are the resulting output units when the "Absolute Densities" option is selected). Separate KDEs were performed for each of the bandwidths estimated for the case data.

Dual KDEs were performed on the case and population data from the non-weighted and weighted versions of the artificial dataset. The fixed distance dual KDEs were performed using both the *CrimeStat* and *Disease Mapping and Analysis Program* (*DMAP*) software; while, the adaptive distance dual KDEs were performed just in *CrimeStat III*, as *DMAP* only offers fixed distance dual KDEs. As with the single KDEs, separate dual KDEs were performed for each of the bandwidths estimated from the population data. All of the dual KDE results were based on kernels with a uniform distribution. In *CrimeStat* this distribution was selected from the distributions offered (which are the five kernel distributions shown in **Figure 5.4B**), while in *DMAP* this is the only distribution available. "Ratio of densities" was chosen as the output units for the dual KDEs performed in *CrimeStat* because the resulting case to population ratios were similar to *DMAP*'s "clustering rates", and thus, the same color-schemes could be used for both.

DMAP also tests the significance of the clustering rates produced by its dual KDEs by performing Monte Carlo simulations in which the clustering surface is tested against a given number of simulated surfaces (described in detail in the *DMAP* section of **Chapter 4**). "1000" was chosen as the total number of simulations for *DMAP* to run for each of its KDE estimates, so that the resulting simulated areas of significant clustering would have a 99.9% confidence level.

The resulting spatial KDEs were brought into *ArcMap* where they could be displayed and visually interpreted and compared to each other, as well as the pre-defined artificial cluster locations. The single and dual KDEs from *CrimeStat* were already in raster form when they were brought into *ArcMap*. The density estimates associated with the single KDEs of the case data were then color-ramped in red, with the darker shades indicating increased case densities. The case to population ratios from the dual KDEs (also referred to as "clustering rates") were color-ramped in green, with the darker end of the color spectrum indicating higher rates of case clustering.

Unlike *CrimeStat*, *DMAP* does not export its results in a form that can be brought directly into *ArcMap*. Instead, the "RATE" and "SIGNIF" files from each analysis must be converted to dbf format. The dbfs were then brought into *ArcMap* where they were joined to a point-based grid of the study area. The rates and Monte Carlo significance values were then interpolated using *ArcMap*'s Inverse Distance Weighting (IDW) tool, resulting in raster surfaces of the clustering rates and significant clustering areas. The IDW raster surfaces for the clustering rates were color-ramped using the same color scheme as the one used for the dual KDEs from *CrimeStat.* The IDW surfaces for the Monte Carlo simulations were used to extract the boundaries for areas in which the clustering rates had statistically significant *p*-values. The boundaries of areas with a *p*-value of 0.05 were then outlined in red to indicate which areas had statistically significant clustering rates.

5.5 Spatial Scan Statistic Methods

5.5.1 Introduction, Purpose, and Importance of the Techniques in This Category

Spatial scan statistics are one of the most common types of spatial analysis used by spatial epidemiologists to detect the locations of current, past, or even future disease clusters (Kulldorff et al. 2005; Robertson and Nelson 2010). The mechanics used by spatial scan statistics are quite similar to those used by adaptive spatial KDEs, in that both use a "scanning" window that adjusts the size of the window based on the number of point locations (Kulldorff et al. 2005). In fact both the fixed and dual KDEs have been referred to as scan statistics (Cai et al. 2011); and conversely, the scan statistics addressed here can also be thought of as spatial filtering and smoothing techniques. While both analyses use circular kernels to detect clusters, the term "spatial filtering" generally refers to the types of kernel density estimates described in the previous section, whereas, "spatial scan statistic" almost always refers to analyses performed using *SaTScan* software. For this reason the two techniques have been divided into their own respective ESDA categories.

5.5.2 Common Types of Spatial Scan Statistics

There are three main types of spatial scanning statistical analyses provided in *SaTScan*: purely spatial, purely temporal, and space-time analyses (Kulldorff 2010). As their names imply, *purely spatial* analysis takes only the locations (and potentially weights) of the numerator and denominator into account, ignoring any temporal information that may be included in the dataset. Conversely, *purely temporal* analysis ignores the geographic locations, and focuses exclusively on the temporal trends in the data. *Space-Time* analysis offers the best of both worlds by taking both the geographic and temporal information into account. Additionally, *SaTScan* offers both *Retrospective* and *Prospective* versions of the above types of analysis. Because of their predictive nature, *prospective* analyses require temporal data and can therefore only be

performed as *purely temporal* or *space-time* analyses; whereas, *retrospective* analyses can be performed as any of the types (Kulldorff 2010). After selecting what type of analysis is to be performed (i.e. *purely temporal, purely spatial, or space-time*; and *retrospective* or *prospective*), the user must select which probability model the analysis will use. *SaTScan* uses a number of models depending on the type of data being analyzed, such as: the *Poisson model* (both discrete and continuous versions); *Bernoulli model*; *Space-Time Permutation model*; *Multinominal model*; *Ordinal model*; *Exponential model*; or the *normal model* (Kulldorff 2006). A brief description of each of these models, as well as recommendations for when to use each of them was provided in the *SaTScan* section of the previous chapter (see **Table 4.5**, on page 89).

5.5.3 Spatial Scan Statistical Analyses Performed on the Artificial Dataset

Of the eight types of probability models offered by *SaTScan*, only the *discrete Poisson* and *Bernoulli models* are appropriate for the artificial dataset. The *Bernoulli Model* examines presence/absence data and was therefore used only on the non-weighted case and population data, which meant that all of the case and control locations were given a weight of either 1 (indicating presence) or 0 (indicating absence). The *discrete Poisson Model* examines count data and was therefore used only on the weighted case and population data, in which the locations were weighted by the number case or total individuals present at the given location. Given that there is no temporal information associated with the artificial dataset, both models were run as *retrospective purely spatial* analyses (as opposed *purely temporal* or *Space-Time retrospective* and *prospective* analyses). Since *SaTScan* uses scanning windows with constantly changing filter radii, a comparison of different bandwidths was not appropriate. Instead, the results of *SaTScan* analyses based on the following criteria for reporting secondary clusters were compared: the default, No Geographical Overlap (NGO); No Cluster Centers in Other Clusters (NCCOC); No Cluster Centers in More Likely Clusters (NCCMLC); and No Restrictions (NR).

The results of all the analyses were then brought into *ArcMap* where the location, size, and significance of the clusters detected using the *Bernoulli* and *Poisson models* and each of the four options for reporting secondary clusters, could be displayed and compared. All of the results, including the relative risk (RR) and least likelihood ratio (LLR) estimates, for both models were also summarized in table format.

5.6 Spatial Autocorrelation (SA) Methods

5.6.1 Introduction, Purpose, and Importance of the Techniques in This Category

As with the previous ESDA categories, there is also overlap in which types of analysis are included under the "Spatial Autocorrelation" (SA) category. Strictly speaking *Spatial Autocorrelation* is defined as a statistical measure of spatial dependence (the clustering or dispersion of point locations) in which the likelihood that the given spatial distribution could have occurred through chance alone is also calculated (Myint 2010). Based on this definition, almost all of the previously discussed techniques (with the possible exception of those discussed in the Mapping & Visualization ESDA category) could be considered types of SA analysis. However, traditionally the term "Spatial Autocorrelation" has been used to refer to local and global versions (or variations) of the following three statistics: *Geary's C*; *Getis-Ord G*; and *Moran's I* (Goodchild 1986; Getis and Ord 1992; Anselin 1995; Fotheringham 1997; Fotheringham et al. 2000; Levine and Associates 2004; Anselin et al. 2006a; Getis 2010; Myint 2010; Ratcliffe 2010; Hu et al. 2011). All three of these statistics require that the location data being analyzed be in the form of continuous polygons (rather than points) in which each polygon represents the location of a feature with an associated attribute value. Given that none of the previously discussed methods required the point data to be in this form (in fact, most required that it be in the opposite form, as *x*,*y* point locations), only the three "traditional" SA statistics are included under the Spatial Autocorrelation ESDA category.

Spatial autocorrelation (SA) is a statistical measure of the degree of spatial association (also referred to as "spatial dependence") between the geographic locations and attribute values associated with each feature in a given dataset (Goodchild 1988; Fotheringham et al. 2000,2002; de Smith et al. 2009; Griffith 2009; Myint 2010; Fischer and Wang 2011). Spatial patterns in which a feature with a given attribute value (such as a high or low number of individual coral colonies present at that location) are surrounded by features with *similar* attribute values show "positive" spatial autocorrelation (**Figure 5.5A$_1$**; Goodchild 1988; Griffith 2009; Myint 2010). Conversely, when the attribute values associated with neighboring features tend to be more *dissimilar* than the values of features located further away, then the spatial pattern as a whole shows "negative" spatial autocorrelation (**Figure 5.5A$_3$**; Goodchild 1988; Griffith 2009). Last, when the distribution of the attribute values appear to be independent of the geographic location of their associated features, this suggests that the attribute values and feature locations are not spatially autocorrelated (Goodchild 1986) and therefore represent a random spatial distribution (**Figure 5.5A$_2$**).

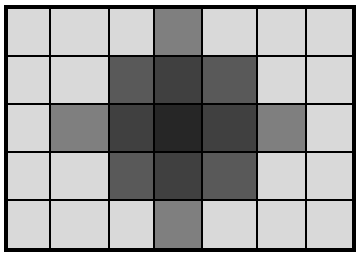
5.6.2 Common Types of SA Analysis

Global measures of spatial autocorrelation are used to identify the presence and absence of clustering, assuming that the results of the analysis apply uniformly to the entire dataset (Anselin 1995). Whereas, local measures are used to find where the clustering (or lack thereof) is occurring. In other words, local SA analyses would detect local differences in the spatial distribution of diseased corals within the study area, while global measures assume these differences do not exist (Fotheringham et al. 2000). General differences between global and local spatial statistics were provided earlier in **Table 5.1** (on page 98).

The following paragraphs describe the basic characteristics of the three most common spatial autocorrelation statistics (*Geary's C*, *Getis-Ord G*, and *Moran's I*). More detailed

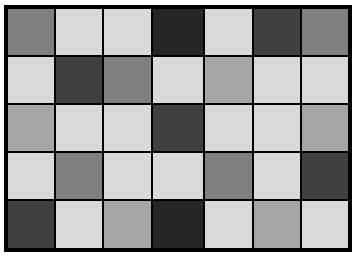
(A) General Types of Spatial Autocorrelation (SA)

(A_1) Positive SA [1,2]



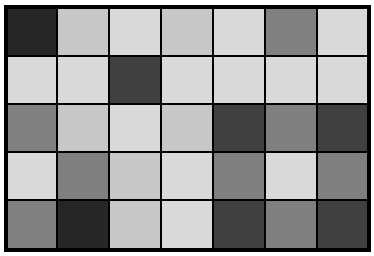
Features with similar attribute values are spatially clustered, representing an aggregated spatial distribution

(A_2) No (Zero) SA [2]



The spatial distribution of features and their associated attribute values appears to be random

(A_3) Negative SA [1,2]



Features are surrounded by features with dissimilar attribute values, representing an anti-clustered (dispersed) distribution.

(B) Specific Types of Positive SA

(B_1) "**Hotspots**" formed by clustered *highs* [3]

48	52	51	85	49	48	52
49	51	90	95	90	49	51
50	85	95	**100**	95	85	85
51	49	90	95	90	51	49
52	48	49	85	51	52	48

Feature with a *high* value is surrounded by features with *similarly high* attribute values, suggesting that the *high* value features are part of a spatial cluster.

Statistically Significant *Hotspots* are referred to as "**High-High**" (**HH**) clusters.

(B_2) "**Cold Spots**" formed by clustered *lows* [3]

48	52	51	15	49	48	52
49	51	10	5	10	49	51
50	15	5	**0**	5	15	50
51	49	10	5	10	51	49
52	48	49	15	51	52	48

Feature with a Low value is surrounded by features with *similarly* Low attribute values, suggesting that the *low* value features are part of a spatial cluster.

Statistically Significant *Cold Spots* are referred to as "**Low-Low**" (**LL**) clusters.

(C) Specific Types of Negative SA

(C_1) "**Spatial Outlier**" with a High feature value

48	52	51	10	49	48	52
49	51	5	0	5	49	51
50	10	0	**100**	0	10	50
51	49	5	0	5	51	49
52	48	49	10	51	52	48

Feature with a *high* attribute value is surrounded by features with dissimilar, *low* attribute values, suggesting that the *high* valued feature is a spatial outlier.

Statistically Significant ($p \leq 0.05$) *high value Spatial Outliers* are referred to as "**High-Low**" (**HL**) outliers.

(C_2) "**Spatial Outlier**" with a Low feature value

48	52	51	90	49	48	52
49	51	95	100	95	49	51
50	90	100	**0**	100	90	50
51	49	95	100	95	51	49
52	48	49	90	51	52	48

Feature with a *low* attribute value is surrounded by features with dissimilar, *high* attribute values, suggesting that the *low* valued feature is a spatial outlier.

Statistically Significant ($p \leq 0.05$) *low value Spatial Outliers* are referred to as "**High-Low**" (**HL**) outliers.

Figure 5.5 Common patterns detected through Spatial Autocorrelation (SA) analyses. Note: superscript numbers indicate that the given diagram is based on concepts derived from the following published figures: [1]Myint (2010)'s Figure 1 on page 2607; [2]Fischer and Wang (2011)'s Figure 2.3 on page 24; and [3]Myint (2010)'s Figure 2 on page 2608.

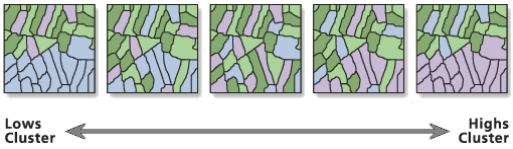
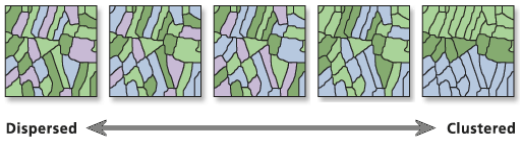
information is provided for the global and local versions of the *Getis-Ord G* and *Moran's I* SA statistics in **Tables 5.2** and **5.3**, respectively. The *Geary's C* was not included in either table because I was unable to find a spatial analysis program that would compute this statistic.

The *Geary's C* statistic estimates the degree of spatial association based on squared difference measurements of the correlation between the attribute values and locations of spatial features in a given dataset (Fischer and Wang 2011). The results of both global and local versions of this statistic are assessed using the *Geary Ratio* (*GR*), which is a quantitative index of spatial autocorrelation. The null hypothesis of complete spatial randomness (CSR) is accepted when the calculated index values are equal to 1 and rejected for $C \neq 1$. Positive SA is indicated by index values between zero and one ($0 < C < 1$), while index values greater than one ($C > 1$) suggest negative SA (Goodchild 1986; Lee and Marion 1994; Fischer and Wang 2011).

The *Getis-Ord G* statistic measures the amount of positive SA and determines whether this clustering is of high or low feature values (**Figure 5.5B**). The global version of the *G* statistic (sometimes referred to as the *General G*) estimates the amount and type of positive SA for the entire study area, while the two local versions of the statistic (G_i and G_i^*) estimate the amount and type of clustering for the each feature located in the study area. Positive SA is characterized by the spatial clustering of features with similar attribute values, resulting in the formation of local "hotspots" and "cold spots" (Fischer and Wang 2011).

Hotspots" occur when features with high attribute values are surrounded by features that have similar high attribute values (**Figure 5.5B$_1$**), and conversely, "cold spots" occur when features with low attribute values surround features that also have low attribute values (**Figure 5.5B$_2$**). As with the global version (*G*), both the G_i and G_i^* statistics are measures of clustering intensities of features with either high or low attribute values. The G_i statistic measures the amount of positive SA for a given location (*i*) by the strength of the correlation between the

Table 5.2 Global Spatial Autocorrelation (SA) Statistics

	***Getis-Ord General G* (*G*)**	***Moran's I* (*I*)**
Description[1,2]	"Measures the degree of clustering for either high values or low values"	"Measures spatial autocorrelation based on feature locations & attribute values"
Conceptual Illustrations[1,2]		
Equation[3,4]	$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j}, \ \forall j \neq i$	$I = \frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}z_i z_j}{\sum_{i=1}^{n} z_i^2}$
Null Hypothesis (*H₀*)	H_0 : Complete Spatial Randomness (CSR), spatial distribution of the feature values is random Accept H_0 when $p \geq 0.05$	Reject H_0 when $p \leq 0.05$
Index Values	The range of General G (*G*) Index values varies Observed *G* > Expected *G* = + *z* : highs are clustered Observed *G* < Expected *G* = – *z* : lows are clustered Observed *G* = Expected *G* : *z* = 0 : random distribution	Moran's Index (*I*) values are usually between ±1 + *I* : indicates a tendency towards clustering – *I* : indicates a tendency towards dispersion *I* ≈ 0 : indicates spatial distribution is random .
Critical Values (*z* – scores)	*z*-scores indicate the clustering type & intensity + *z*-scores : high values cluster together – *z*-scores : low values cluster together *z*-score = 0 : no apparent clustering	*z*-scores indicate the distribution type & intensity + *z*-scores : high &/or low values are clustered – *z*-scores : high &/or low values are dispersed *z*-score = 0 : no apparent clustering
***ArcMap* Results Output**	No Spatial Output. The Observed & Expected Index Values, *z*-score, & *p*-value represent the overall spatial distribution of the feature values within the study area	

Note: superscript numbers indicate that the given material was taken directly from the following sources: [1]ESRI (2009h); [2]ESRI (2009i); [3]ESRI (2009j); and [4]ESRI (2009a). The rest of the information shown in the above table was based on the following references: Goodchild (1986); Waller and Gotway (2004); de Smith et al. (2009); and Myint (2010).

locations and attribute values for the features surrounding the location *i*; whereas, the G_i^* statistic measures the correlation strength between the attribute value of location *i* and the locations and values of the surrounding features (Fischer and Wang 2011). In other words the attribute value and location of the feature of interest (*i*) are included in the clustering estimate of the G_i^* but not the G_i statistic.

The Moran's I SA statistic estimates the degree of spatial association based on cross-product measurements of the correlation between the attribute values and locations of spatial features in a given dataset (Fischer and Wang 2011). The key difference between Moran's I and

Table 5.3 Local Spatial Autocorrelation (SA) Statistics

	Getis-Ord Local G (G_i^*)	***Local Moran's I*** (I_i)
Description	Identifies <u>where</u> features with a given attribute value are surrounded by features with <u>similar</u> attribute values suggesting that these features are part of a cluster. Clusters made up of features with high attribute values are referred to as *Hot Spots*; while clusters made up of low values are referred to as *Cold Spots*. The G_i^* statistic also estimates the statistical significance associated with each of these *hot* & *cold* spots. Significant *Hot Spots* are referred to as *High-High* (*HH*) clusters & significant *Cold Spots* are referred to as *Low-Low* (*LL*) clusters.	Identifies <u>where</u> *Hot* & *Cold Spots* are located; as well as, the locations of *Spatial Outliers*. "Spatial outlier" refers to a feature surrounded by features with <u>dissimilar</u> attribute values. The I_i statistic also estimates the statistical significance associated with each of the spatial clusters & spatial outliers. Significant clusters are referred to as *HH* & *LL*. Significant spatial outliers with high attribute values (surrounded by features with low attributes values) are referred to as *High-Low* (*HL*) outliers, while low value outliers are referred to as *Low-High* (*LH*) outliers.
Conceptual Illustration[1,2]		
Equation[3,4]	$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}}$	$I_i = \frac{x_i - \bar{X}}{S_i^2}\sum_{j=1,j\neq i}^{n} w_{i,j}(x_i - \bar{X})$
Null hypothesis (H_{0_i})	H_{0_i} : Complete Spatial Randomness (CSR), spatial distribution of the feature values is random. The *p*-values are used to determine whether to Accept or Reject the H_{0_i} feature by feature. Accept H_{0_i} when $p \geq 0.05$ and Reject H_{0_i} when $p \leq 0.05$	
Index Values	The G_i^* statistic is a *z*-score, So for any given feature the computed G_i^* value will be the same as the computed *z*-score value	$I_i > 0$: surrounding attribute values are <u>Similar</u> $I_i < 0$: surrounding attribute values are <u>Dissimilar</u> $I_i = 0$: surrounding attribute values are <u>Random</u>
Critical Values (***z* – scores**)	$\lvert z \rvert$ indicates the degree of spatial clustering $z > 0$: <u>high</u> values <u>cluster</u> together $z < 0$: <u>low</u> values <u>cluster</u> together $z = 0$: no apparent clustering in attribute values	z indicates the type &$\lvert z \rvert$ degree of spatial dependence $z > 0$: <u>high</u> or <u>low</u> values cluster together $z < 0$: <u>high</u> or <u>low</u> valued <u>spatial outlier</u> $z = 0$: spatial distribution of values is <u>Random</u>
Significance (***p* – values**)	Represents the statistical significance of spatial clustering of values	Represents the statistical significance of the computed Index values (I_i)
***ArcMap* Results Output**	Spatial Output of Local Clusters G_i^* *z*-scores (GiZScore) & *p*-values (GiPValue) can be used to create choropleth maps that show both the locations & statistical significance of clusters	Spatial Output of Local Clusters & Spatial Outliers I_i Index values (LMiIndex), *z*-scores (LMiZScore), *p*-values (LMiPValue), & Cluster Types (COType) can be used to create choropleth maps showing the locations & significance of clusters & spatial outliers

Note: superscript numbers indicate that the given material was taken directly from the following sources: [1]ESRI (2009b); [2]ESRI (2009c); [3]ESRI (2009e); and [4]ESRI (2009g). The rest of the information shown in the above table was based on the following references: Anselin (1995); Fortin and Dale (2005); and de Smith et al. (2009).

Getis-Ord G statistics is that the *I* statistic also estimates the amount of negative SA, whereas the *G* statistic just deals with the positive SA. The global version of the *I* statistic can therefore be used to determine whether the overall spatial distribution of a given dataset is more clustered or dispersed than would be expected through chance alone. Whereas, the local version of the statistic (I_i) can be used to detect the location and significance of features with spatially clustered values (**Figure 5.5B**), as well as, features with spatially dispersed values (**Figure 5.5C**).

5.6.3 SA Analyses Performed on the Artificial Dataset

The two most common types of Global SA analysis, the *Getis-Ord General G* and *Moran's I*, were performed on the Artificial dataset using the "High/Low Clustering (Getis-Ord General G" and "Spatial Autocorrelation (Moran's I)" tools from the "Analyzing Patterns" toolset provided in *ArcMap*'s "Spatial Statistics" toolbox. Both the *Getis-Ord General G* and the *Moran's I* SA analyses were performed on weighted versions of the artificial case, control, population (case + control), and prevalence (estimated as the number of case individuals divided by the total number of individuals present at each location) data separately. The results from these analyses were then displayed in a table where they could be compared. Local versions of both statistics, G_i^* and *Local Moran's I* (I_i) were also performed on the same data using the "Hot Spot Analysis (Getis-Ord Gi*)" and "Cluster and Outlier Analysis: Anselin's Local Moran's I" tools from the "Mapping Clusters" toolset, which is also provided in *ArcMap's* Spatial Statistics toolbox. The cluster types, index values, *z*-scores, and *p*-value estimates from each of the local analyses were then compared using choropleth maps of their values. For comparative purposes *Univariate* and *Bivariate Local Moran's I* analyses were also performed using the *OpenGeoDa* software. *Univariate Local Moran's I* analyses were performed separately on the artificial case, control, and prevalence data using the "Univariate LISA" tool located under the "Space" menu option in *OpenGeoDa*. A *Bivariate Local Moran's I* analysis was also performed on the

artificial case and population data using the "Multivariate LISA" tool, which is also located in the Space menu. A spatial weights matrix based on queen contiguity was used for all *Local Moran's I* analyses performed in *OpenGeoDa*. The resulting analysis output was joined to the original Thiessien polygon shapefile, which was then brought into *ArcMap* so that the index values, cluster types, and cluster significance estimates could be displayed as choropleth maps.

All of the SA analyses (both global and local) were performed on weighted versions of the artificial dataset in which the point locations were displayed as a continuous polygonal surface using projected versions of the Thiessen polygons created in section 5.2.3. The null hypothesis of complete spatial randomness (CSR) was either accepted or rejected based on the individual results calculated by each of the SA statistics. The results for each statistic were assessed using the information provided in **Tables 5.2** and **5.3**, and statistical characteristics associated with the standard normal distribution (**Figure 5.6**).

5.7 Spatial Regression Methods

The following section provides a brief overview of spatial regression analytical methods. Spatial regression is designed to evaluate and model the spatial relationship between two or more attributes associated with a minimum of several hundred-feature locations (ESRI 2009d). The artificial dataset used to assess the other spatial methods in this chapter, was a relatively small dataset designed to represent a relatively small geographic area. Given that the artificial dataset only contained 44 case locations, which would have been the variable of interest, spatial regression was not appropriate. Consequently a more comprehensive review of this ESDA category has not been included as part of this dissertation.

5.7.1 Spatial Regression

Spatial regression is often the final step in ESDA because it goes beyond just visualization and cluster detection, allowing the relationships between different spatial variables

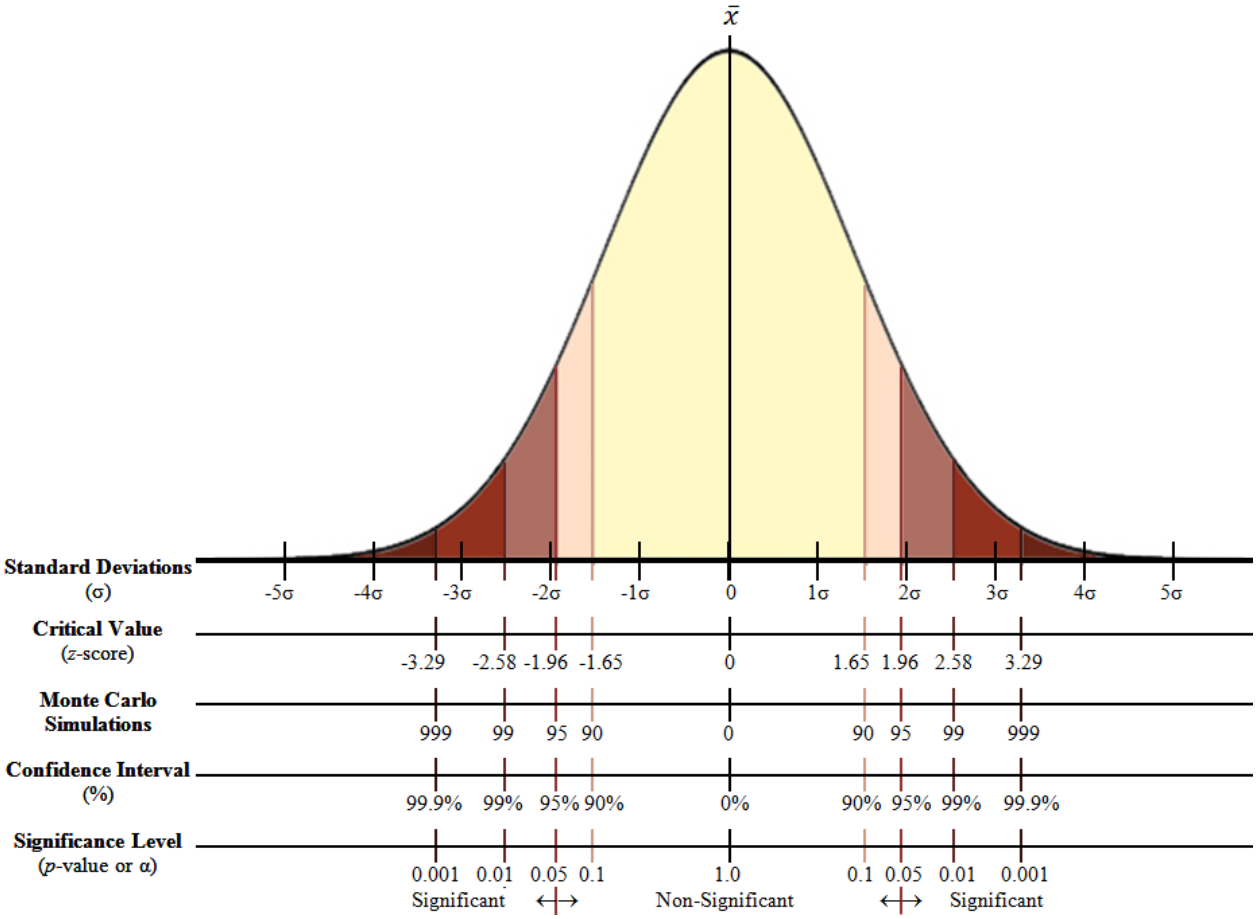


Figure 5.6 The *p*-values and *z*-scores calculated during the various types of spatial autocorrelation (SA) analyses performed are all based on the Standard Normal Distribution (shown above). The significance of the SA results can be determined by locating the calculated *p*-values and *z*-scores on the above figure. Note: this figure was adapted from Figure 6.4 on page 111 of Spatz and Johnston (1976) and Appendix C10 on page 219 of Ebdon (1985).

to be modeled in order to better explain which factors (independent variables) have the most influence on the spatial nature of the dependent variable (Anselin 2005). Well specified regression models can be used to explain the phenomena of interest, test hypotheses, and potentially even predict future outcomes (ESRI 2009f; Rosenshein et al. 2011). Regression techniques are used to model the linear relationship between a dependent variable and one or more independent variables (**Figure 5.7A**; Charlton and Fotheringham 2009). The "dependent" variable (y), which is also referred to as "response" variable or the "Regressand", is the variable or process which is trying to be understood (Charlton and Fotheringham 2009; ESRI 2009f).

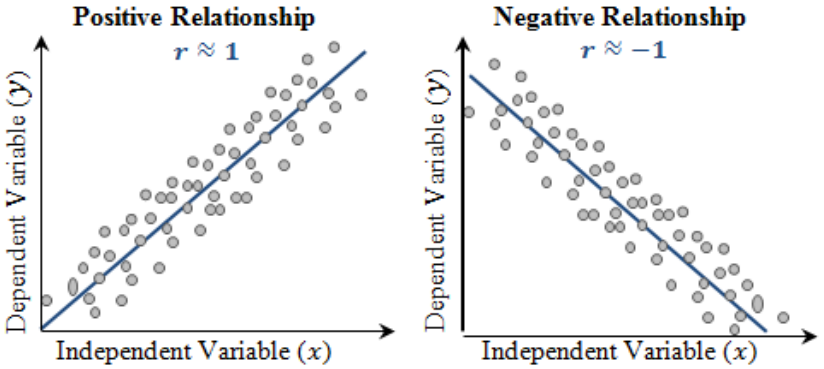
The "independent" variable (x), also known as the "predictor" or "exploratory" variable , or as "Regressors" when referring to more than one x, is the variable(s) which are being used to try to help explain the dependent variable (Charlton and Fotheringham 2009; ESRI 2009f). In the case of coral disease, the disease of interest would be the dependent variable, and the independent variables would be factors that might help model the disease of interest, such as: other coral disease data; environmental data (salinity, temperature, depth, acidity, etc.); or anthropogenic stressors.

In addition to being the "best known" of the regression techniques (ESRI 2009f), Ordinary Least Squares (OLS) regression (**Figure 5.7C**) provides the foundation for Geographically Weighted Regression (GWR) analysis (**Figure 5.7D**), and therefore serves as the starting point for any type of spatial regression analysis (Tu and Xia 2008; Charlton and Fotheringham 2009; ESRI 2009f; Rosenshein et al. 2011). OLS is used to create a global model of the variable or process of interest, such as: whether or not the spatial distribution of a given coral disease is related to specific anthropogenic stressors, and if so, which stressors appear to be more influential than others. The results of OLS are used to build a well specified GWR and provides guidance on selecting the key exploratory variables. GWR is then used to model this relationship at the local level by fitting the OLS regression equation to each individual feature in the dataset. In this way, OLS and GWR methods can provide powerful, and a statistically robust way of analyzing linear relationships spatially. See **Figure 5.7** for additional information.

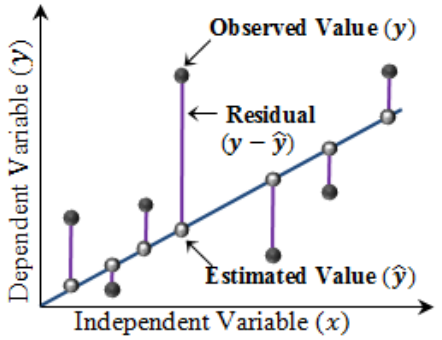
5.8 Summarizing the ESDA Methodology

Table 5.4 provides a summary of the six Exploratory Spatial Data Analysis (ESDA) categories, the types of analysis generally included under each category, and examples of the types of information that can be obtained from each ESDA category and analysis type. The results of the analyses described in this table are presented in the following chapter.

(A) Common Linear Relationships modeled using Regression analyses



(B) Residuals



(C) Ordinary Least Squares (OLS) Regression

Provides a global model of the relationship between 2 or more variables

$$\hat{y}_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i$$

- The relationship between the dependent (y) & independent variables (x) is the same (homogeneous) throughout the study area.
- Any geographic spatial variation present must be confined to the error term
- Variance is constant
- The model's residuals are uncorrelated (no autocorrealation)
- No spatial output

(D) Geographically Weighted Regression (GWR)

Models the local spatial variability of the relationship between 2 or more variables

$$\hat{y}_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

- Allows spatial variation (heterogeneity) between the dependent (y) & independent (x) variables by estimating the parameters at individual locations (u_i, v_i).
- Observation i's parameter estimates are influenced more by nearby observations than by those further away
- Improves the reliability of relationships by reducing the amount of spatial autocorrelation
- Regression equation is calibrated independently for each observation i, resulting in separate parameter estimates & regression statistics for each location (spatial output)

where…

$\hat{y}_i$ is the estimated value of the dependent variable for observation i

x_{ik} is the value of the k^{th} independent variable for i

(u_i, v_i) is the geographic location (coordinates) of observation i

β_0 is the intercept

β_k is the parameter estimate for variable k

ε_i is the error term

Figure 5.7 Additional information related to spatial regression. Note: the above figure was adapted from the following sources: Figure 5.1 on page 186 of Devore and Peck (2005); Mennis (2006); and Charlton and Fotheringham (2009).

Table 5.4 A summary of the different types of Exploratory Spatial Data Analyses (ESDA) discussed in this chapter and examples of the types of information that can be attained from each of them.

Type of Spatial Analysis	Spatial Information Attained
1. Mapping & Visualizing Data • Mapping Point Locations using points & polygons • Scaling Point Symbols &/or colors to visualize intensity	**Visualizing Spatial Distributions** • Visualizing the spatial distribution of data locations • Visualizing the spatial distribution of data density (or intensity of an attribute)
2. Point Pattern Analysis o Centrographic Statistics • "Mean Center" estimates • *Median Center* (*MdnCntr*) • *Minimum Convex Polygons* (*MCP*) • Standard Distance & Deviation Estimates o Distance Statistics • *Nearest Neighbor Analysis* (*Nna*) • *Ripley's K* (*K*)	**Describe the General Spatial Distribution of the Data** o Demonstrate the location & spatial distribution of point patterns • Identifying the central focal point of the points • Useful when outliers are influencing the mean center • Simplest method for estimating the Home Range of an animal • Estimate the general distribution of the data around a central focal point o Test hypotheses regarding the spatial distribution of points • Examine spatial dependence (clustering or dispersion) at a given scale • How spatial dependence changes with distance & scales of measurement
3. Spatial Filtering and Smoothing • Single Kernel Density Estimates (KDE) • Dual KDEs • *DMAP*'s Dual KDE with Monte Carlo simulations	**The Presence, Degree, & <u>Location</u> of Clusters** • Density, Intensity, and Probability estimates • Prevalence, Odds Ratios, & Relative Risk Estimates • All of the above plus Significant Clustering Areas
4. Spatial Scan Statistics • Spatial & Temporal Scan Statistics	**Scan Statistics are used to detect Outbreaks through the Cluster Analysis** • Cluster Size, Significance, Relative Risk, Changes with time
5. Spatial Autocorrelation (SA) o Global SA • *Getis-Ord General G* • *Moran's I* o Local SA • *Getis-Ord* G_i^* • *Local Moran's* I_i	**Whether or Not Clustering is Present** o Whether or not Spatial Autocorrelation (SA) is present region-wide • Measures the degree of clustering for either "high" or "low" values • Measures the amount of SA based on feature locations & attribute values o Where local SA is present • Identifies where "high" or "low" values cluster spatially • Identifies the locations of high & low clusters, as well as spatial outliers
6. Spatial Regression • *Ordinary Least Squares (OLS) Regression* • *Geographically Weighted Regression* (*GWR*)	**Performs local regression analyses without assuming spatial homogeneity** • OLS results output is used to build the GWR model • Assesses spatial heterogeneity between independent & dependent variables

Chapter 6. Results of the ESDA of the Artificial Dataset

"Characterizing reef sites where disease distribution and prevalence, and the frequency, intensity and distribution of epizootic events, is different from other reefs might provide important information to better understand the spatial/temporal dynamics of coral and octocoral diseases in the region." – Weil and Croquer (2009)

6.1 Mapping and Visualization Results

The results of the point and polygon-based visualization techniques used for the non-weighted and weighted versions of the artificial dataset are shown in **Figure 6.1**. The non-weighted and weighted versions of the artificial case and control data were depicted using the same shades of red and green for both the point and polygon-based visualization techniques.

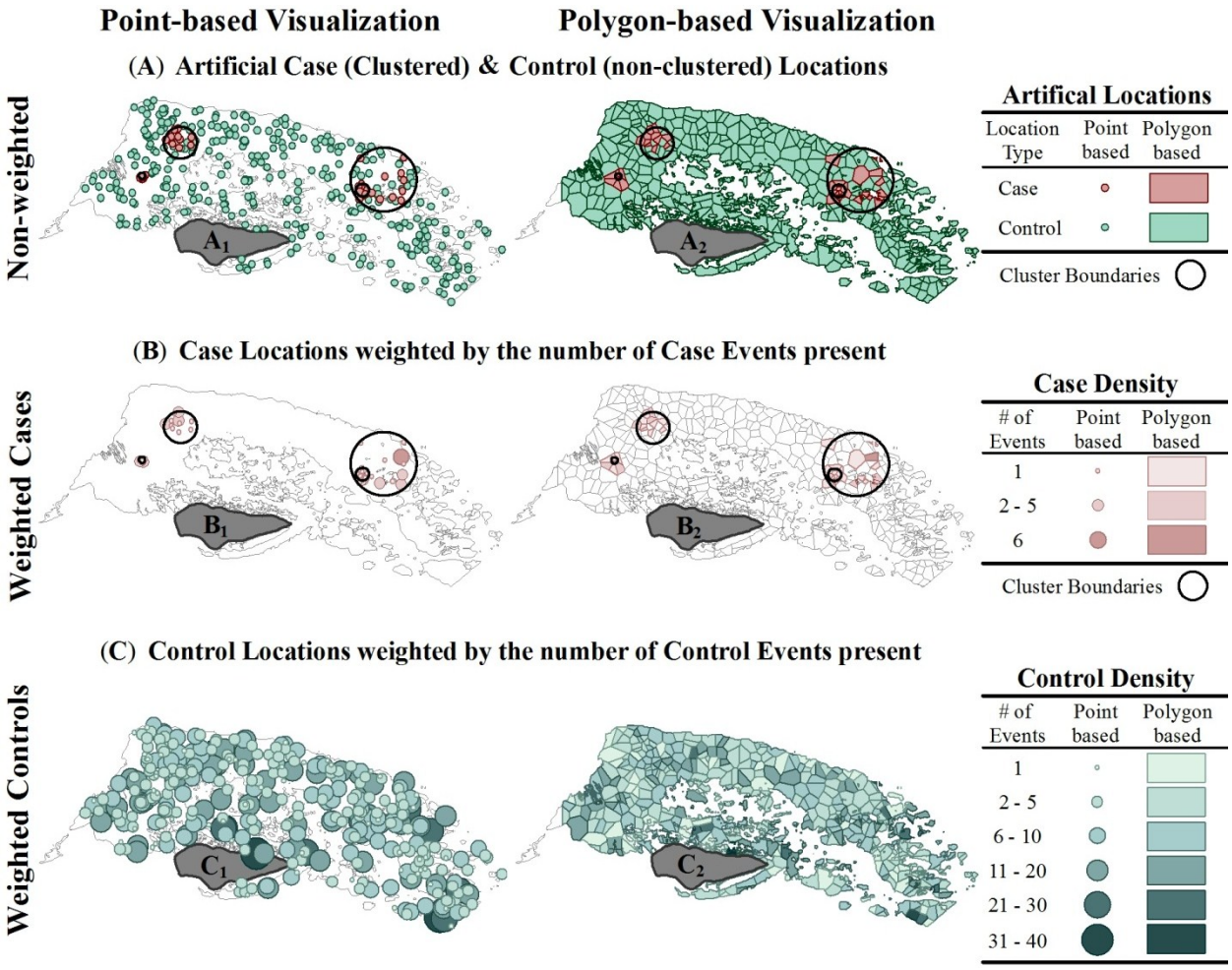


Figure 6.1 Different techniques for visualizing the Artificial Cluster Dataset

The locations of the artificial clusters were more accurately depicted using the point-based method (**Figure 6.1A$_1$** and **B$_1$**) because the area of the Thiessen polygons associated with the case data often extended past the artificial cluster boundaries (**Figure 6.1A$_2$** and **B$_2$**). However, the polygon-based visualization method was more effective for displaying the case and control densities (**Figure 6.1B$_2$** and **C$_2$**, respectively) because their non-overlapping boundaries made the variations in color easier to distinguish; while the overlapping point symbols were distracting and harder to interpret (**Figure 6.1B$_1$** and **C$_1$**).

In cases where only the spatial distribution and attributes associated with the case locations are of interest, it may be useful to create a "prevalence map" in which the number of case individuals is divided by the total number of individuals present at a given location (**Figure 6.2**). Prevalence maps offer a way of visualizing case density while also accounting for the underlying population density; however, all information regarding the population densities of non-case locations is lost.

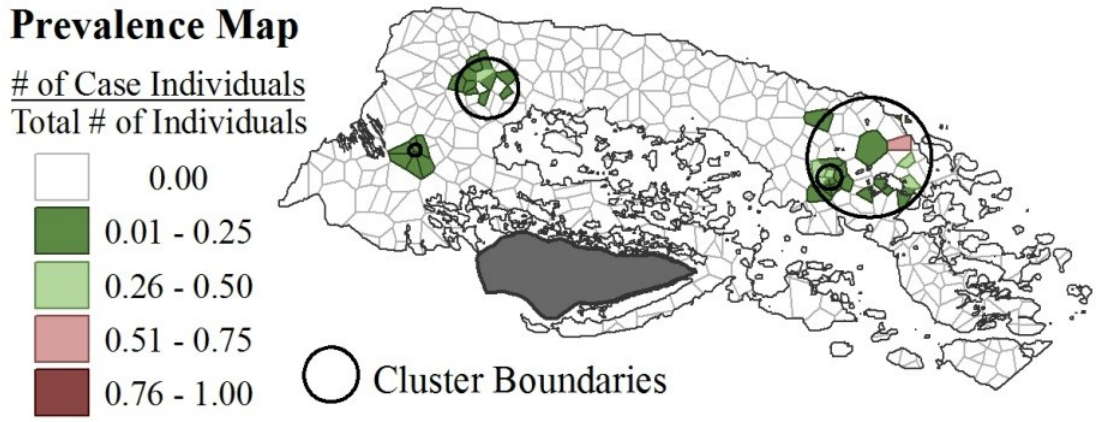


Figure 6.2 Prevalence Map, based on the weighted Artificial Case and Population data, using the polygon-based visualization technique.

For this reason, point-based visualization techniques are recommended for portraying data locations (non-weighted data); while, polygon-based visualization techniques are recommended for portraying the attributes values associated with the feature locations (weighted data).

6.2 Point Pattern Analysis (PPA) Results

6.2.1 Results of the Centrographic Statistical PPA of the Artificial Case Locations

The results of the centrographic statistical analyses performed on the artificial case locations are summarized in **Figure 6.3**. There was a slight difference in the anticipated location of the mean center, based on which of the three individual mean center statistics was used (**Figure 6.3A** close-up). However, the difference between the locations was too small (<5*m*) to be visually discernible without zooming in (**Figure 6.3A**). The locations of the *Median Center* (**Figure 6.3B**) and the *Center of Minimum Distance* (**Figure 6.3A**) for the case data were located near the general location of the predicted mean center, with the *Median Center* falling in between the mean centers and the *center of minimum distance* (**Figure 6.3G**).

In addition to encompassing all of the case locations, which was expected based on how *minimum convex polygons* (*MCP*) are defined and created; the *MCP* also encompassed the entire area of three case clusters and the majority of the 4th cluster based on the cluster boundaries (**Figure 6.3C**). The results of the standard distance and deviation statistical analyses are shown in Figure **6.3D-F**. The *Standard Distance Deviation* (**Figure 6.3D**) encompassed the entire cluster area for both the 100*m* and 250*m* clusters, but only included two case locations for the 500*m* cluster and none of the case locations for the 50*m* cluster. The polygon representing the *Standard Deviation of the X and Y coordinates* (**Figure 6.3E**) encompassed all of the cases in the 100*m* cluster, a few of the cases from both the 250*m* and 500*m* clusters, and none of the cases from the 50*m* cluster. The *standard deviational ellipse* based on 1 standard deviation also encompassed all of the cases from the 100*m* cluster while also including more cases from both the 250*m* and 500*m* clusters and bordered the location of the 50*m* cluster (the 1x ellipse in **Figure 6.3F**).

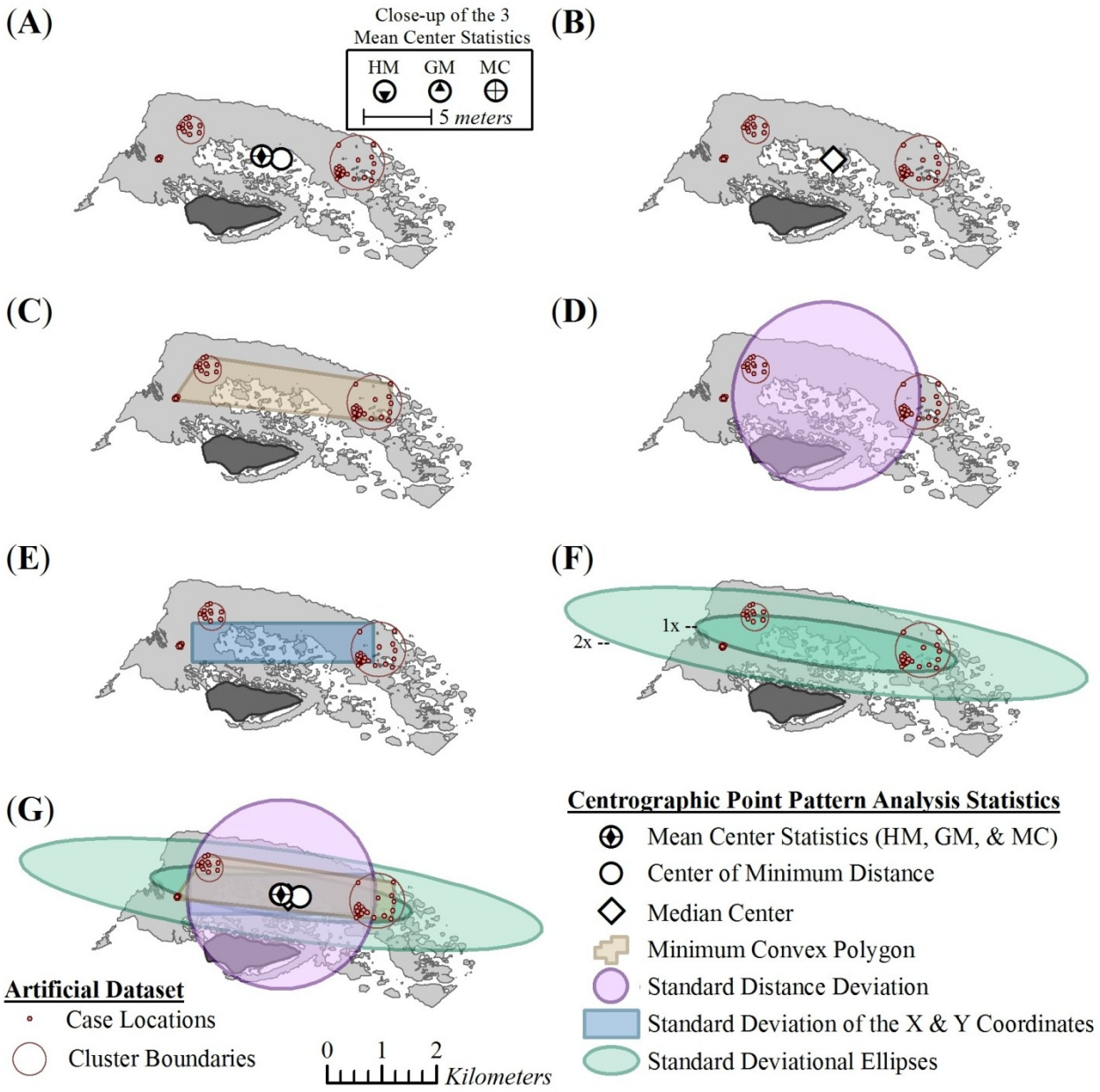


Figure 6.3 Centrographic Statistical Point Pattern Analyses (PPA) performed on the case locations from the non-weighted artificial dataset using the "Spatial Distribution Tools" in *CrimeStat III*. The results of the mean and median center statistical analyses are shown in **A-B**, the *Minimum Convex Polygon* (referred to in *CrimeStat* as the "convex hull") is depicted in **C**, the results of the individual standard distance and deviation statistical analyses are shown in **D-F**, and last **G** depicts the results of all of the centrographic statistics. Specifically, (**A**) shows the location of the *Center of Minimum Distance* and the general location of the mean center based on the specific locations of the *Harmonic Mean* (*HM*), the *Geometric Mean* (*GM*), and the *Mean Center* (*MC*), which are shown in the inset; (**B**) shows the location of the *Median Center*; and (**D-F**) shows the area encompassed by the *Standard Distance Deviation* (**D**), the *Standard Deviation of the X and Y Coordinates* (**E**), and the *Standard Deviational Ellipses* (**F**) based on 1 standard deviation (1x) and 2 standard deviations (2x).

6.2.2 Results from the *Nearest Neighbor Analysis* (*NNA*) of the Case Locations

The results of the *Nearest Neighbor Analysis* (*NNA*) of the artificial case locations based on the first order of *K* are given in **Table 6.1**. Statistically significant ($p \leq 0.0001$) spatial clustering (NNI < 1) was found between case locations and their first nearest neighbor ($K = 1$), with a mean nearest neighbor distance of 73.53*m*.

Table 6.1 Results of *Nearest Neighbor Analysis* (*NNA*) of the Artificial Case Locations, based on first order ($K = 1$) Nearest Neighbor Index (NNI) values.

Mean Nearest Neighbor Distance	73.53*m*	Nearest Neighbor Index (NNI)	0.3485
Standard Deviation of Nearest Neighbor Distance	85.00*m*	Standard Error	16.63*m*
Minimum Distance	5.74*m*	Test Statistic (*z*)	- 8.2675
Maximum Distance	3,994.45*m*	*p* – value (one tail)	≤ 0.0001
Mean Dispersed Distance	453.43*m*	*p* – value (two tail)	≤ 0.0001
Mean Random Distance based on the size of the Study Area*	210.99*m*		

*The size of the study area was based on the area of the surveyed benthic habitat which was 7,834,831.06m^2

Plotting the NNI values against the 44 orders of *K*-Nearest Neighbors revealed that the case locations tended to be spatially clustered (NNI < 1) for the first 21 orders of *K* (the 21 nearest neighbors), but spatially dispersed (NNI > 1) for *NNA* based on higher orders of *K* (**Figure 6.4**). The strongest clustering (NNI < 0.5) was detected for the first 10 orders of *K*; increasing to NNI ≈ 0.75 for the next 10 orders of *K*.

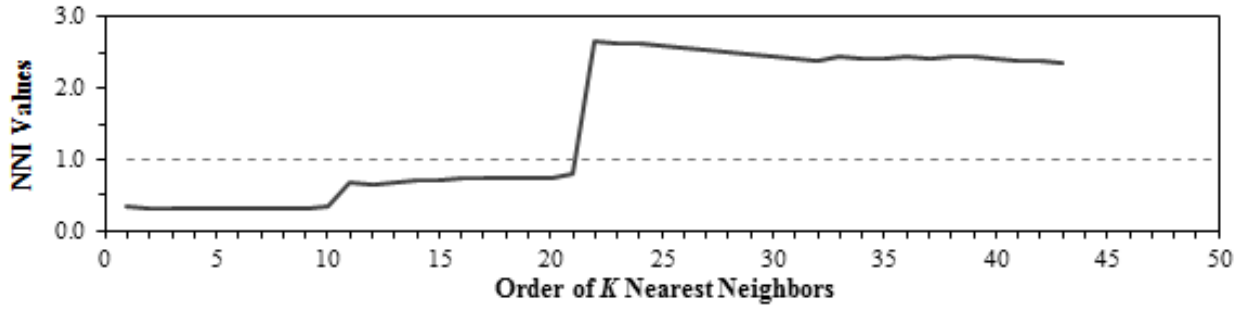


Figure 6.4 Graph of the Nearest Neighbor Index (NNI) values for the *Nearest Neighbor Analysis* (*NNA*) of the Artificial Case locations for the 44 orders of *K* Nearest Neighbors tested. NNI < 1 indicates Spatial Clustering and NNI > 1 indicates Spatial Dispersion.

This makes sense, given that each of the artificial clusters has 11 case locations associated with it, and therefore, the 10 nearest neighbors of any given case location would be the locations of the 10 other cases associated with the given case cluster. Thus, the strongest clustering should be detected for the 10 nearest neighbors, as orders above 10 would be testing the spatial dependence of cases located in different clusters. Given that the cluster with the 100*m* radius is located within the boundary of the cluster with the 500*m* radius, and the cluster with the 50*m* radius is located near the cluster with the 250*m* radius. It would also follow that spatial clustering is likely to be detected between the cases located in one cluster and the cases in the nearest other cluster, which explains why clustering was detected between any given case location and the 21 case locations closest to it (i.e. its 21 nearest neighbors).

6.2.3 Results from *Ripley's K* Analysis of the Artificial Case and Population Data

The individual results of the *Ripley's K* analyses performed on the non-weighted and weighted versions of the artificial case and population data are shown in **Figure 6.5**. The benchmark for evaluating complete spatial randomness (CSR) is depicted by the solid grey line, representing the *Expected K* values. *Observed K* values that fall above this line are considered spatially clustered, while *Observed K* values falling below this line are considered spatially dispersed. Based on these analyses, the non-weighted case locations were found to be spatially clustered to a statistically significant extent ($p = 0.01$) for distances < 550*m*, significantly dispersed between 550*m* and 600*m*, and dispersed (but not to a statistically significant extent) at distances greater than 600*m* (**Figure 6.5A**).

Ripley's K analyses of the weighted case locations (**Figure 6.5B**) also detected clustering at distances < 550*m* and dispersion at distances > 550m; however, given that the *Observed K* values fell between the upper and lower confidence intervals the observed clustering and dispersion were not statistically significant. *Ripley's K* plots of the underlying artificial

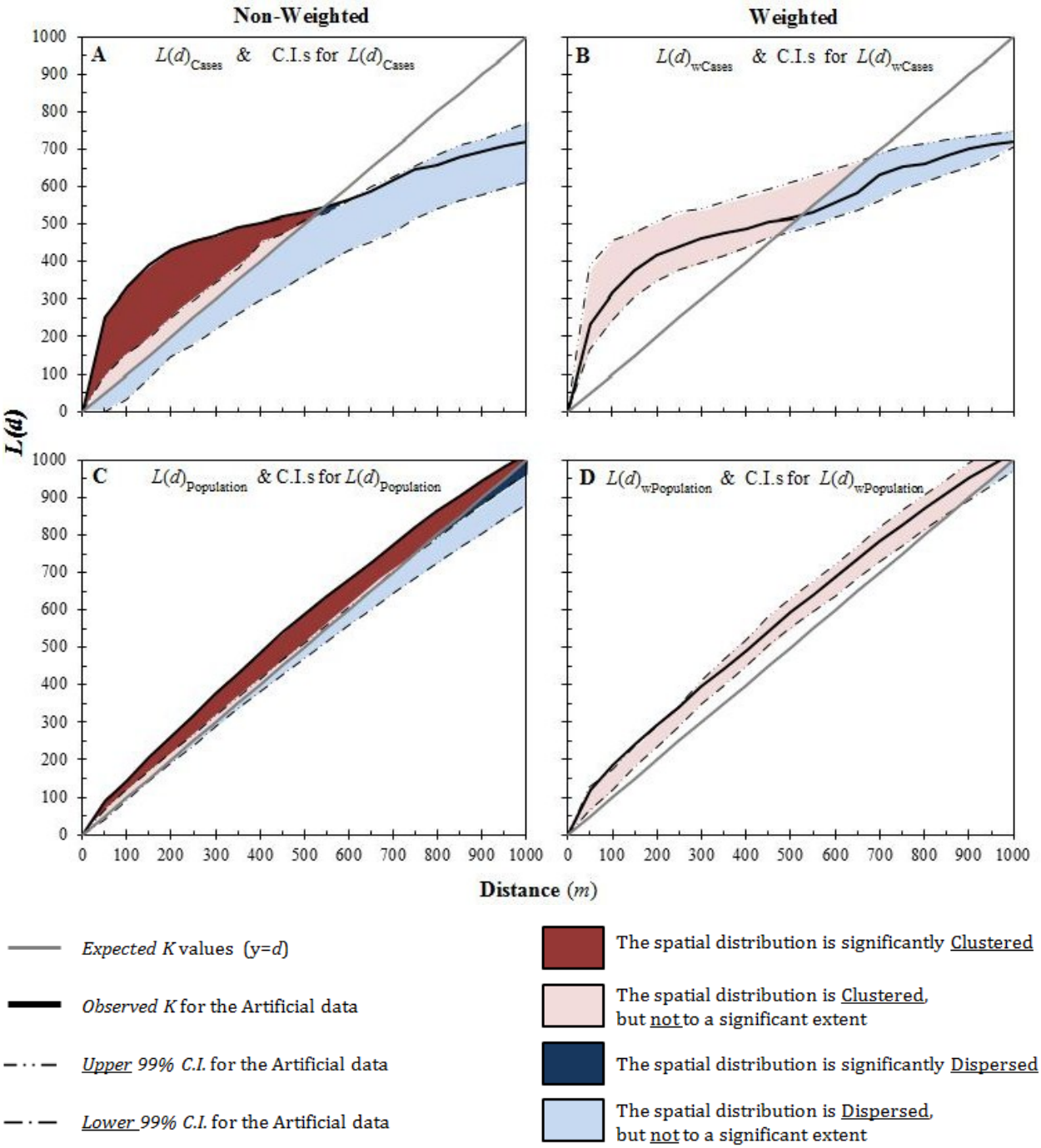


Figure 6.5 *Ripley's K* plots comparing the spatial distributions of the non-weighted and weighted versions of the artificial case and population datasets. Note, the spatial distribution is considered statistically significant when the *observed K* values fall outside of the 99% confidence envelope. For this reason the area between the *observed K* values and the upper 99% confidence interval (C.I.) is considered to have a statistically significant spatial distribution, with significant clustering occurring above the *expected K* values ($y=d$) and significant dispersion occurring below $y=d$.

population data found spatial clustering at all distances tested (**Figure 6.5C** and **6.5D**). This clustering was found to be significant for all distances (0*m* – 1,000*m*) based on the non-weighted locations of the artificial population data (**Figure 6.5C**); but not significant when the locations were weighted by the number of individuals within them (**Figure 6.5D**).

Figure 6.6 shows the same information as **Figure 6.5**. However, the values in **Figure 6.6** have been normalized by subtracting the expected values (d) from the observed values ($L(d)$). The benchmark for evaluating CSR for normalized values then becomes $y = 0$, as opposed to the prior, non-normalized CSR benchmark of $y = d$. Thus, even though the results are the same as those described above the hyperbolic nature of the plots has been removed, making the resulting graphs much more expressive. For this reason, only normalized data were used to create the plots of the remaining *Ripley's K* analyses.

The difference function (D) was used to compare the spatial distribution of the case data to the spatial distribution of the underlying population (**Figure 6.7**). The spatial distribution of case locations were found to be significantly more clustered than the clustered spatial distribution of the underlying population locations at distances less than 400*m*, and significantly more dispersed than the dispersed spatial distribution of the distances greater than 400*m* and less than 500*m*; however, case locations were found to be more clustered at distances less than 450*m* and more dispersed than the underlying population at distances greater than 500*m* (**Figure 6.7A**). No significant difference was detected between the spatial distributions of the case and population locations at distances greater than 450*m*.

A similar trend was detected when the difference function was used to compare the spatial distribution of the weighted case locations to the spatial distribution of the weighted underlying population locations (**Figure 6.7B**). When the case locations were weighted by the number of individual cases present at each location, the spatial distribution of cases was found to

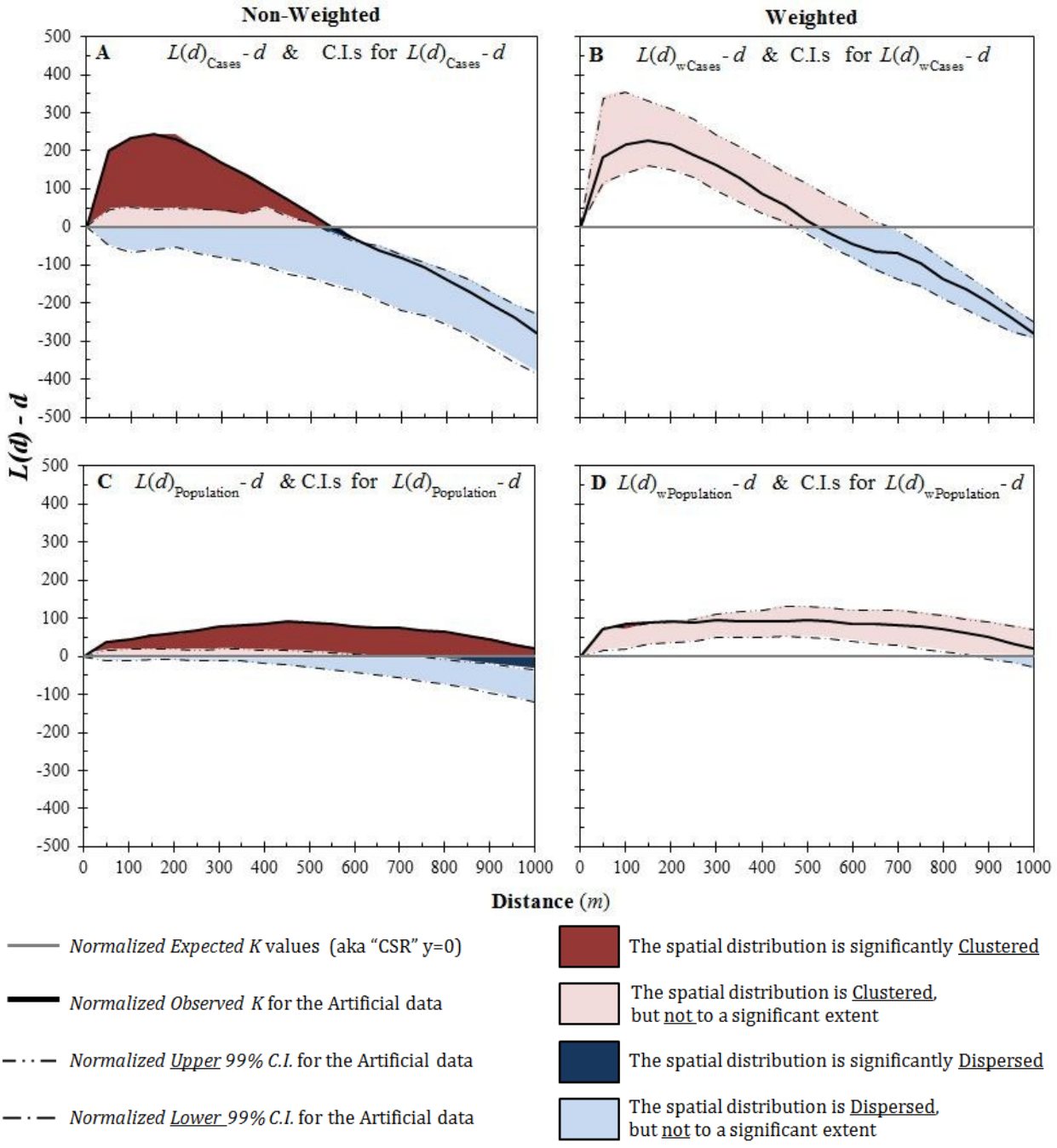


Figure 6.6 *Normalized Ripley's K* plots depicting the same information as shown in **Figure 6.5**, but here the *Ripley's K* results have been normalized. Notice how the areas with significant spatial distributions are slightly more discernible in the normalized plots above (**A** and **C**), than they were in the non-normalized plots shown in **Figure 6.5 A**,**C**.

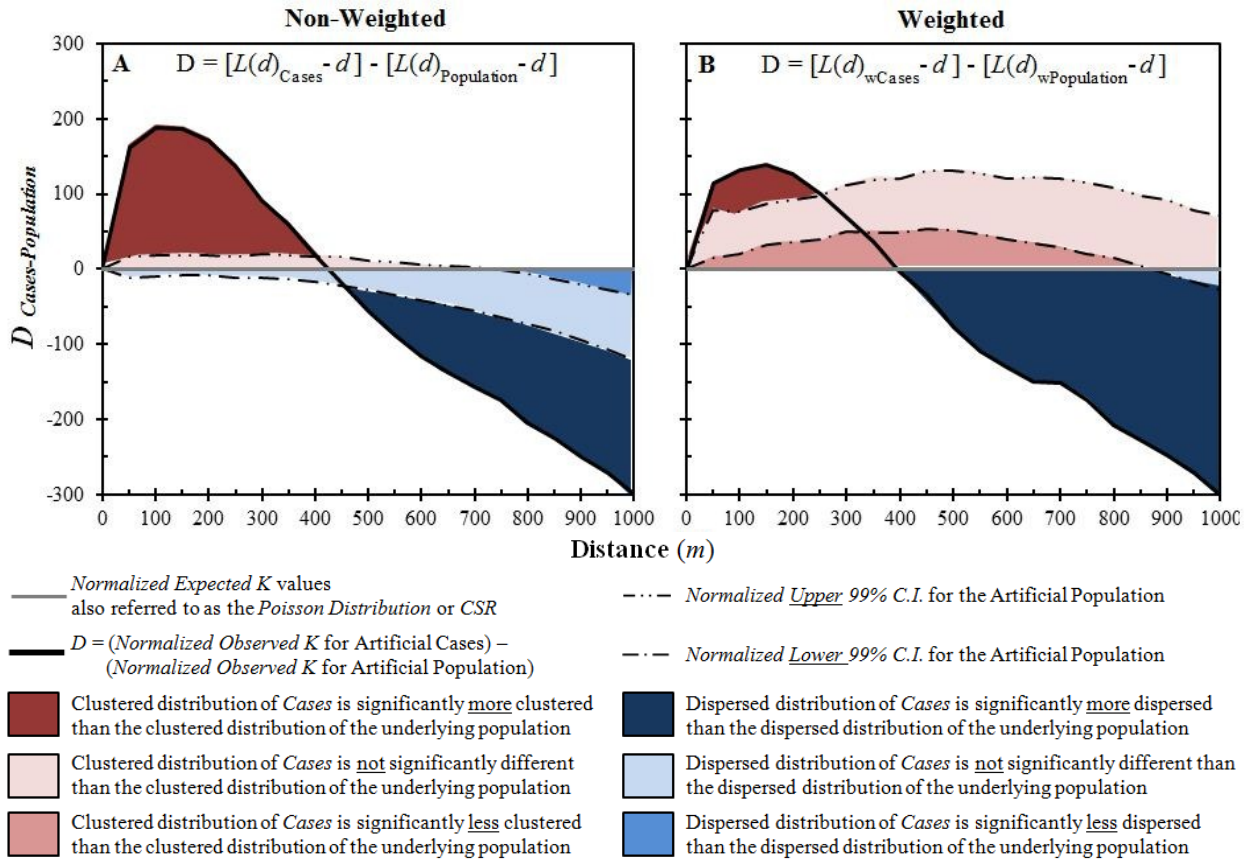


Figure 6.7 Results of the *Ripley's K* analyses performed on the artificial dataset. *Normalized Ripley's K* plots were used to assess the spatial distribution of the artificially clustered points (i.e. case locations) compared to the spatial distribution of all of the artificial data points (i.e. case and control locations). (**A**) *Ripley's K* analyses performed on the non-weighted version of the artificial dataset (i.e. point locations). (**B**) *Ripley's K* analyses performed on weighted versions of the artificial dataset (i.e. point densities).

be clustered at distances less than 400*m* and dispersed at distances greater than 400*m*. The spatial distribution of the weighted case locations was more aggregated than the aggregated spatial distribution of the underlying population at distances less than 350*m*, and significantly more aggregated at distances less than 250*m*. Conversely, the spatial aggregation detected for weighted case locations between 350 and 400*m* was more dispersed than the spatial clustering detected in the weighted locations of the underlying population. For distances greater than 400*m*, the spatial distribution of the weighted cases locations was significantly more dispersed

than the spatial distribution of the underlying population, which was clustered between 400 and 900m and dispersed at distances greater than 900m.

Figure 6.8, shows the graphical test of the null hypothesis that artificial point locations weighted by the number of individuals within them are not significantly more clustered or dispersed than the underlying spatial distribution based on the point locations alone. This hypothesis was rejected for the artificially clustered (case) data at distances less than 550m because the weighted observed K were greater than the upper CI for the non-weighted observed K indicating that case locations weighted by the number of individual case events within them were significantly more clustered than their locations alone would suggest (**Figure 6.8A**).

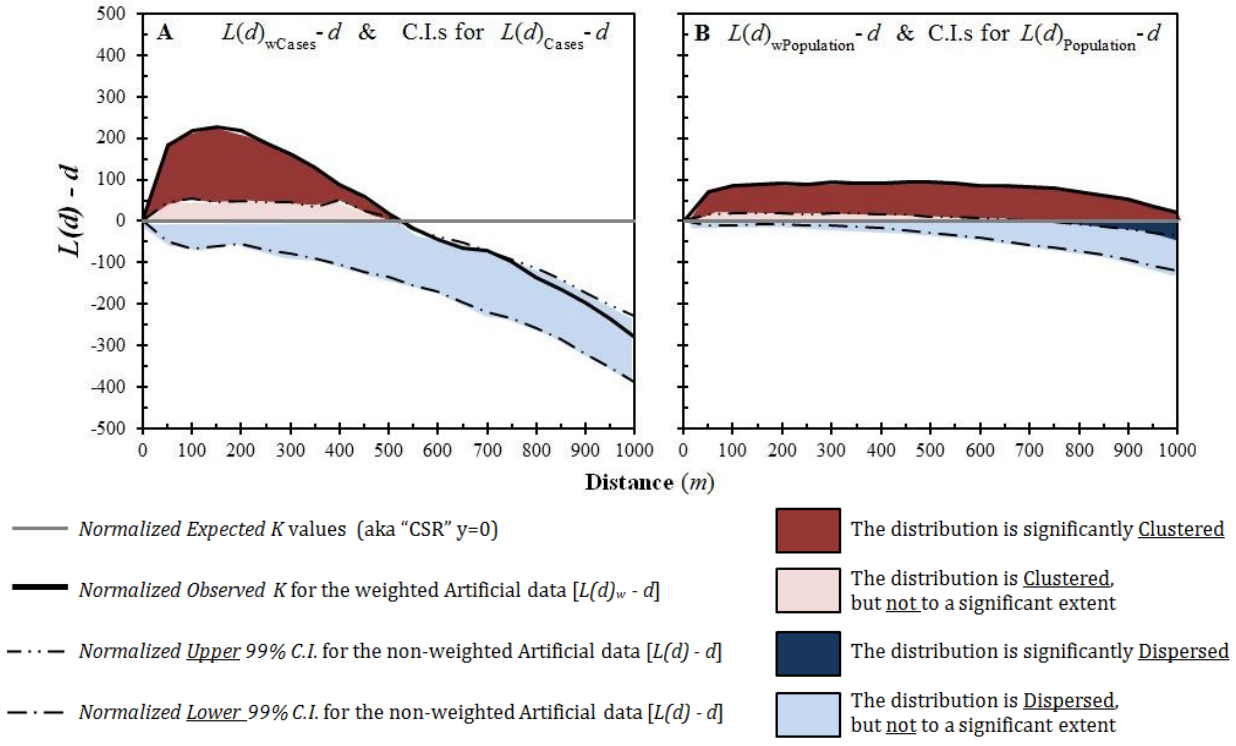


Figure 6.8 A graphical representation of the test of the null hypothesis (h_0) that the spatial distribution of the artificial locations weighted by the number of artificial events occurring within them would not be significantly more clustered or dispersed than the underlying spatial distribution based on the artificial locations alone (i.e. the non-weighted artificial data). In order for the h_0 to be accepted the *Observed K* based on the weighted data (thick black line) must fall within the upper and lower 99% Confidence Intervals (C.I.s, depicted as thin dashed lines) estimated using the non-weighted artificial data. (**A**) The h_0 was rejected at distances <550m and accepted at distances >550m for the artificially clustered data. (**B**) The h_0 was rejected at all of the distances tested for the underlying artificial population.

However, the null hypothesis was accepted when the case data were examined at distances greater than 550*m*, as the *observed K* for the weighted case locations was within the upper and lower CI for the *observed K* of the non-weighted case locations, indicating that at these distance scales, the spatial dispersion of the weighted case locations was not statistically significant from the dispersed distribution of the non-weighted case locations (see **Figure 6.8A**). This hypothesis was rejected for the underlying population for all of the distance scales tested because the weighted *observed K* was above the upper CI for the non-weighted population data, indicating that the point locations weighted by the number of individuals within them were, in fact, significantly more clustered than the spatial distribution of the locations alone (see **Figure 6.8B**).

Figure 6.9 shows a graphical test of the null hypothesis that locations weighted by the number of individuals within them would not be more clustered or dispersed than they would be by chance alone. This hypothesis was accepted for both the artificial case data (**Figure 6.9A**) and the population (**Figure 6.9B**) because the *Observed K* based on the non-weighted location data fell within the CI envelope based on the weighted *Observed Ks*.

Overall, the *Ripley's K* results provided new (and important) information about how the different distance scales affected the estimated spatial dependence (clustering or dispersion) of both the non-weighted and weighted artificial case and population data.

6.2.4 Summary of the PPA Results

Overall, the centrographic statistical PPA techniques were poor predictors of the artificial case and cluster locations; whereas, the distance-based statistical PPA techniques generally offered more insight into the spatial nature of the artificial data. Of the two distance statistical PPA techniques tested, *Nearest Neighbor Analysis* (*NNA*) of the artificial case locations and the *Ripley's K* analysis of the weighted and non-weighted artificial case and population data, the

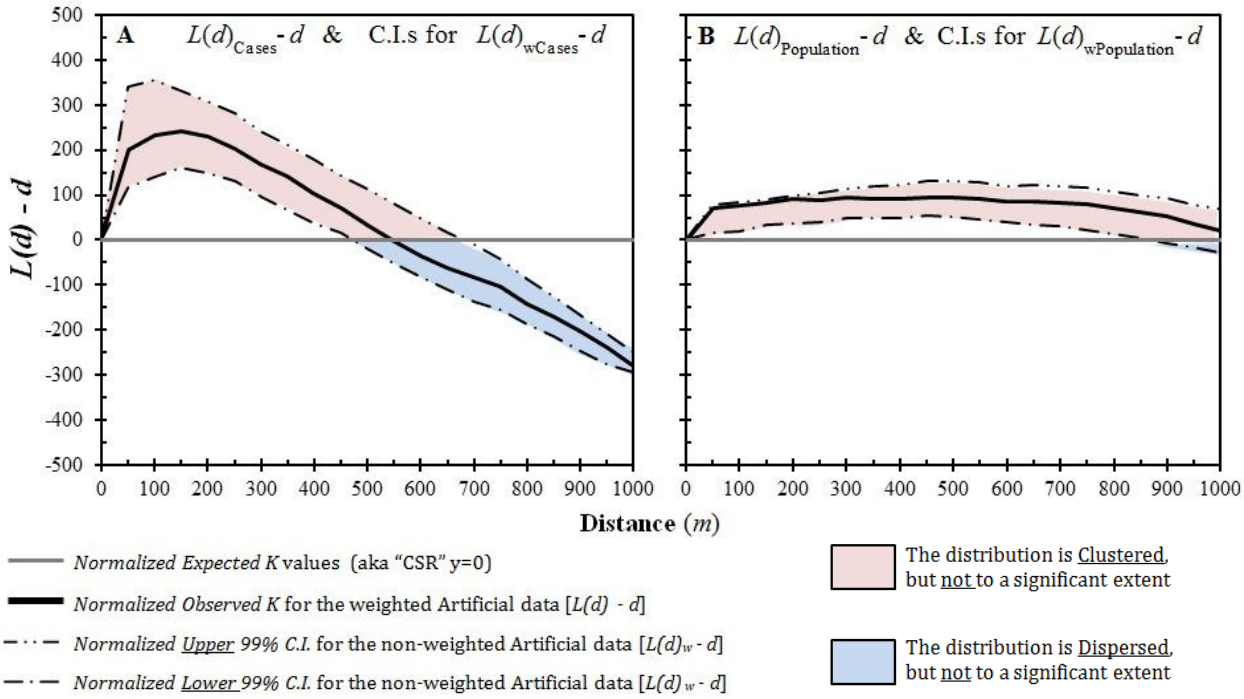


Figure 6.9 A graphical representation of the test of the null hypothesis that the spatial distribution of the weighted artificial data would not be more clustered or dispersed than would be expected through chance alone. This hypothesis was accepted for both (**A**) artificial case data and the (**B**) underlying artificial population because the *Observed K* values (think black line) based on the non-weighted data falls within the 99% Confidence Intervals (C.I.s, depicted as thin dashed lines) based on the *Observed K* estimated using the weighted artificial data.

Ripley's K analyses were preferred over the *NNA*. This is because, while the graphical results of the *NNA* analyses could be explained by the artificial case locations and cluster boundaries, it is unclear how useful these results would be without *a priori* knowledge of the spatial nature of the data being tested.

The *Ripley's K* results provided new information about how the different distance scales affected the estimated spatial dependence (clustering or dispersion) of both the non-weighted and weighted artificial case and population data. Theoretically, the results of the *Ripley's K* analyses could be used to help select the most appropriate bandwidth (filter radius) for fixed distance kernel analyses of the artificial dataset. Additionally, numerous hypotheses regarding the spatial distribution of both the artificial case and population data were able to be tested through the use

of the difference function or by simply plotting different combinations of the *Observed K* and confidence interval results. The major drawback of the *Ripley's K* statistic is that it provides no spatial output of where the clustering is occurring. For this reason, the *Ripley's K* statistic should be used in conjunction with other ESDA techniques that provide spatial output, such as: kernel-based spatial filtering and smoothing analyses; spatial scan statistics; local spatial autocorrelation; or spatial regression analyses.

6.3 Spatial Filtering and Smoothing Results

6.3.1 Spatial Parameter Estimation Results

6.3.1.1 Estimated Grid Cell Resolutions

Of the 121 DMAP analyses performed on the Artificial Cluster dataset during the Visual Calibration Method, only the 62 combinations in which the grid cell resolution was less than or equal to the filter radius completed successfully (**Figure 6.10**; note, due to space constraints only a subset of the 62 combinations are displayed in this figure). Based on a visual examination of these results, the combination of a $50m^2$ grid cell resolution and $100m$ filter radius was found to most accurately detect and represent the artificially clustered points (see **Figure 6.10** inset). Chainey and Ratcliffe (2005) estimation method resulted in a cell size of $21.64m^2$, which was smaller than the size of the surveyed transects and therefore inappropriate for this study. For this reason, all subsequent spatial filtering analyses were based on a grid with $50m^2$ cells, and the $100m$ filter radius was used as the Visual Calibration method's bandwidth (h_{VC}) in the subsequent kernel density estimate (KDE) analyses.

6.3.1.2 Estimated Fixed Distance Bandwidths

The ideal criterion values, and their respective bandwidths, were identified by graphing the results of the *AICc*, *GCV*, and *LSCV* bandwidth selection criterion methods based on the denominator data for the artificial dataset (**Figure 6.11**). According to the *Corrected Akaike's*

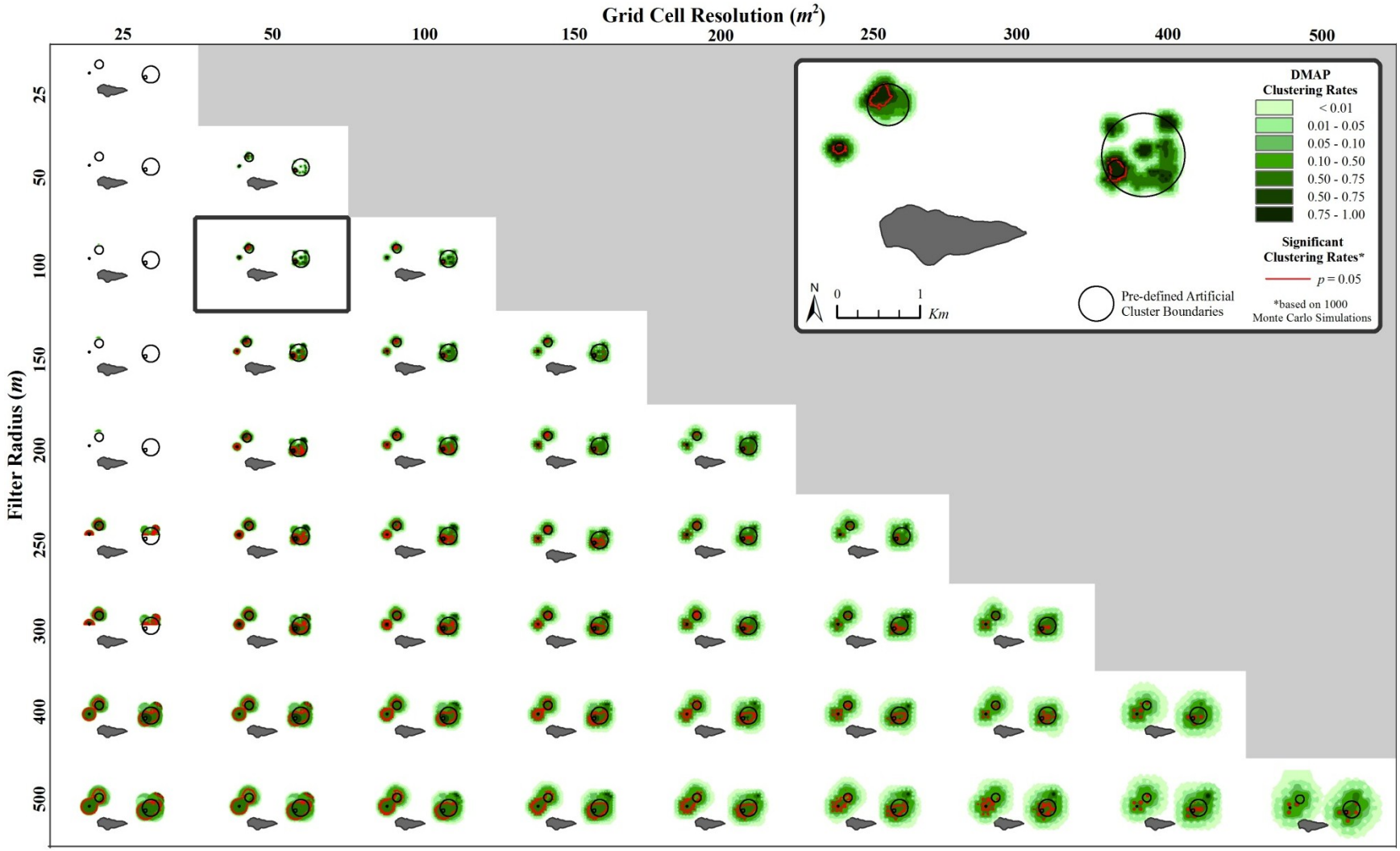


Figure 6.10 Calibrating the *Disease Mapping and Analysis Program* (*DMAP*) using the Artificial Cluster Dataset. The inset depicts the *DMAP* analysis that most accurately detected the four artificial clusters that were on the following spatial parameters: a 50m^2 grid cell resolution, and 100m filter radius. Note: The results from the DMAP analysis using Filter Radii and Grid Cell combinations of 350 and 450m are not included in the above figure due to space constraints.

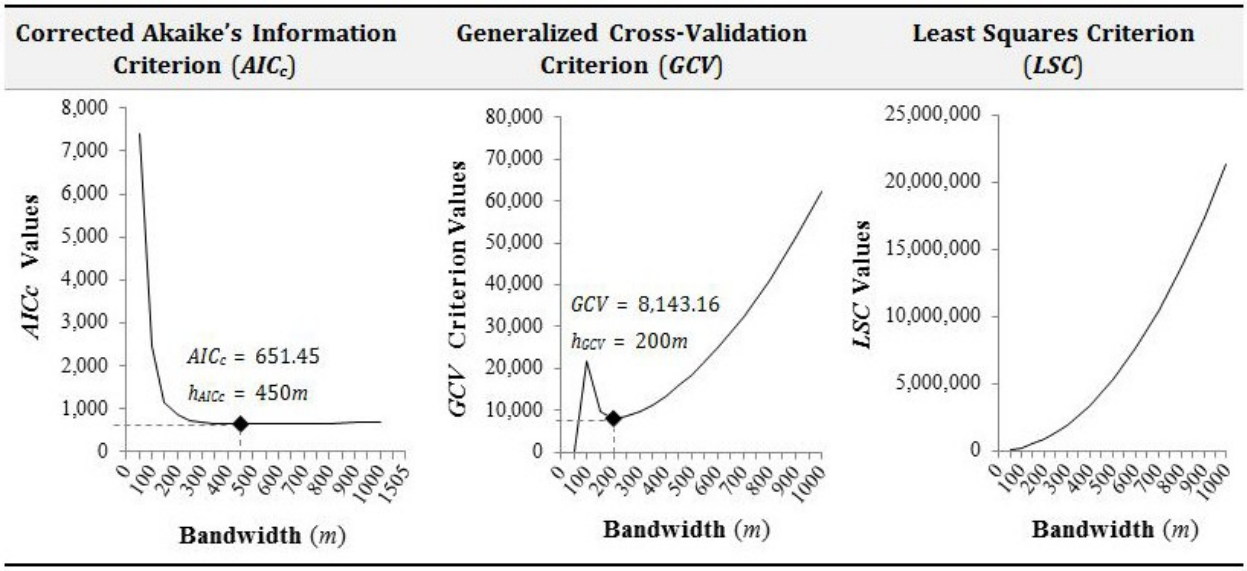


Figure 6.11 The results of the three bandwidth selection criteria methods based on the population data from the artificial dataset. The ideal bandwidth (*h*) is the one with the lowest criterion value, depicted as diamond-shaped symbol. A bandwidth could not be obtained using the Least Squares Criterion (*LSC*) method, as the lowest *LSC* value indicates an optimum bandwidth of 0*m*.

Information Criterion (*AICc*) method, the lowest possible *AICc* value for the artificial dataset was 651.45, which results in an optimal bandwidth of 450*m*. Though, bandwidths greater than 300*m* caused little change in the *AICc* value. The *Generalized Cross-Validation Criterion* (*GCV*) method found the lowest possible *GCV* value for the artificial dataset to be 8143.16, resulting in an optimal bandwidth of 200*m*. The *Least Squares Criterion* (*LSC*) method could not detect an optimal bandwidth, because a minimum *LSC* value could not be found. This was likely a result of the *wrap around effect*, a commonly encountered problem with the *LSC* when it is used in optimal bandwidth selections the calibration wraps itself around the data points.

The results for all the bandwidth calculations based on the numerator (case locations) and denominator (all locations) data for the artificial dataset are summarized in **Table 6.2**. The same bandwidth was selected by the *Biased Cross-Validation* (*BCV2*) and the *Least Squares Cross Validation* (*LSCV*) criterion methods resulting in estimated bandwidths of h_{BCV2} and $h_{LSCV} = 65.7m$

Table 6.2 A comparison of 12 different calibration criteria used to select the most appropriate bandwidth size.

		Numerator data (Artificially Clustered Points)	Denominator data (All Artificial Points)
Corrected Akaike's Information Criterion (AIC_c)	h_{AIC_c} = the h that minimizes the AIC_c	—	$450.00m$
Default search radius in ArcView's (AV) Kernel Density	$h_{AV} = [\min(x, y)] \,/\, 30$	$38.67m$	$106.83m$
Biased Cross Validation (BCV)	h_{BCV2} = h minimizing the $AMISE$ by minimizing the $BCV2(h)$	$65.73m$	$100.10m$
Bailey and Gatrell's (BG) h	$h_{BG} = [0.68n^{-0.2}][\sqrt{A}]$	$892.96m$	$581.72m$
Generalized Cross-Validation Criterion (GCV)	h_{GCV} = h that minimizes the GCV	—	$200.00m$
Least Squares Criterion (LSC)	h_{LSC} = the h that minimizes the LSC	—	—
Least Squares Cross Validation ($LSCV$)	h_{LSCV} = h that minimizes the $MISE$ by minimizing the $CV(h)$	$65.73m$	$100.10m$
Maximal Smoothing (max) Bandwidth	$h_{max} = [(1.144)(\sigma)]/\sqrt[5]{n}$	$923.42m$	$628.81m$
Nearest Neighbor Analysis (NNA) Bandwidth	$h_{NNA} = \sum_{i=1}^{n}\left(\frac{min(D_{ij})}{n}\right)$ estimated for 20 orders of K	K_{01}: $73.65m$ K_{20}: $785.03m$ K_{mean}: $413.95m$	K_{01}: $81.61m$ K_{20}: $460.21m$ K_{mean}: $307.52m$
Optimized (opt) Bandwidth	$h_{opt} = [2/3n]^{\frac{1}{4}}\sigma$	$603.64m$	$369.29m$
Reference (ref) Bandwidth	$h_{ref} = n^{-1/6}\sqrt{(var_x + var_y)/2}$	$641.19m$	$473.37m$
Visual Calibration (VC) using the Artificial dataset	h_{VC}	—	$100.00m$

where…

$\boldsymbol{A}$ is *Study Area*, which is the area of the surveyed benthic habitat. ($\boldsymbol{A} = 7{,}834{,}831.06m^2$)

$\boldsymbol{AMISE}$ is the *Asymptotic Mean Integrated Square Error*, which is a large sample approximation of the $MISE$

$\boldsymbol{h}$ is the size of the *bandwidth* (i.e. the filter radius) measured in meters

$\boldsymbol{MISE}$ is the *Mean Integrated Square Error*

$\boldsymbol{n}$ is the *sample size* which is calculated as the total number of data points ($\boldsymbol{n}_{numerator} = 44$ & $\boldsymbol{n}_{denominator} = 375$)

$\boldsymbol{\sigma}$ is *sigma* (also known as the *standard distance*), which is the estimated *standard deviation of the probability estimate. CrimeStat's* "standard distance deviation" tool calculates *sigma* as $\boldsymbol{\sigma}_{numerator} = 1720.54m$ & $\boldsymbol{\sigma}_{denominator} = 1798.45m$

$\hat{\boldsymbol{\sigma}}$ is *sigma hat,* which is the estimated standard deviation of the error term

$\boldsymbol{tr(S)}$ is the *trace* of the hat matrix (S) which is a function of the bandwidth

v_1 is the *effective number of parameters in the model*, calculated as : $v_1 = \mathrm{tr}(S)$

$var_{x,y}$ is the mean variance in the x and y co-ordinates, respectively

$\boldsymbol{y_i}$ is the value of the *dependent variable* at location i

$\hat{\boldsymbol{y}}_i$ is the *fitted value* (aka. Estimated, Expected, or Predicted value) of y_i

$$\boldsymbol{AIC_c} = 2n\log_e(\hat{\sigma}) + n\log_e(2\pi) + n\left\{\frac{n+tr(S)}{n-2-tr(S)}\right\}$$

$$\boldsymbol{BCV2(h)} = \frac{1}{4\pi nh^2(n-1)} + \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(D_{ij}^2 - 8D_{ij} + 8)e^{-D_{ij}^2/2}}{8(n-1)(n-2)h^2\pi}$$

$$\boldsymbol{CV(h)} = \frac{1}{\pi nh^2} + \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\left(\frac{1}{4\pi}e^{-D_{ij}/4} - \frac{1}{\pi}e^{-D_{ij}/2}\right)}{n^2h^2}$$

$$\boldsymbol{D_{ij}} = \left[\frac{(X_i - X_j)}{h}\right]^2 = \frac{(X_{i(x)} - X_{j(x)})^2 + (X_{i(y)} - X_{j(y)})^2}{h^2}$$

$$\boldsymbol{GCV} = \frac{n\sum_{i=1}^{n}[y_i - \hat{y}_i(h)]^2}{(n - v_1)^2}$$

$$\boldsymbol{LSC} = \sum_{i=1}^{n}[y_i - \hat{y}_i(h)]^2$$

for the numerator data, and h_{BCV2} and h_{LSCV} = 100.1m for the denominator data. The bandwidths based on the *Nearest Neighbor Analysis* (*NNA*) method resulted in a wide range of values based on the 20 orders of *K*. For this reason the average bandwidth ($h_{NNA,k_{mean}}$) for the 20 orders of *K* was calculated for both the numerator and denominator. Four of the bandwidth estimation methods (the visual calibration method and the *AICc*, *GCV*, and *LSC* regression-based criteria) required the use of denominator data and therefore could not be used to generate bandwidths based solely on numerator data.

6.3.1.3 Estimated Adaptive Distance Bandwidths

The estimated adaptive bandwidth for the single KDEs performed on the non-weighted and weighted versions of the artificial case data were $h_{case} \leq 22$, which is 50% of the total number of case locations (*n* = 44). The estimated adaptive bandwidth used for the dual KDEs performed on the non-weighted and weighted versions of the case and population data were also calculated as 50% of the artificial population locations (*n* = 375), resulting in an adaptive bandwidth of $h_{population} \leq 187$.

6.3.2 Spatial Filtering and Smoothing Results

6.3.2.1 Fixed Distance Single Kernel Density Estimates (KDEs) of the Artificial Case Data

The results of the bandwidth comparisons for the single kernel density estimates (KDEs) performed on non-weighted and weighted versions of the artificial case data are shown in **Figure 6.12**. A positive correlation was found between the size of the filter radius and the estimated area of the case density estimates for both the non-weighted and weighted versions of the artificial dataset. KDEs of the artificially clustered points using bandwidths less than 100*m* appeared to accurately predict the case locations and general areas of the 4 clusters. Conversely, KDEs based on bandwidths greater than 100*m* not only captured the entire clustering area (as opposed to just the case locations), but their estimates extended far beyond the cluster boundaries

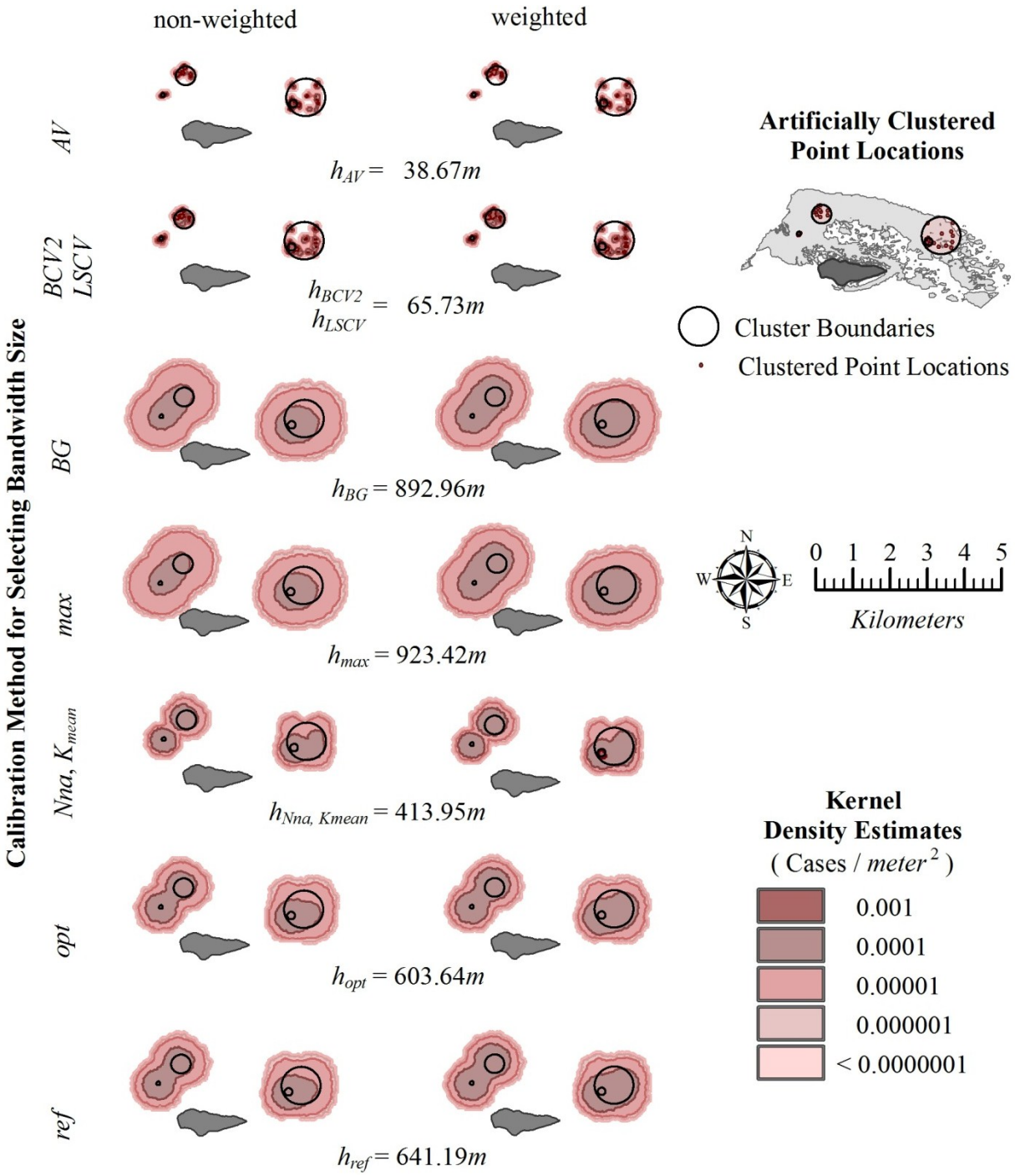


Figure 6.12 *Kernel Density Estimates* (*KDE*) of the weighted and non-weighted versions of the case data. Single *KDE* were performed in *CrimeStat III* using Quartic probability function and fixed distance interpolation. Both the weighted and non-weighted versions of the case data in the artificial dataset was analyzed using 8 of the 12 bandwidth calculation methods, excluding the visual calibration (*VC*) and the 3 regression-based selection criteria methods (*AICc*, *GCV*, and *LSC*) because each required the denominator data in order to complete the computation. The above *KDE* estimates are depicted using the "Relative Density" setting in *CrimeStat*, which divided the absolute density of the case (numerator) data by the area of the grid cells ($50m^2$) resulting in case density estimates per m^2.

providing over-estimates of the clustering area. For any given bandwidth, the overall area in which case density estimates were found for the non-weighted and weighted data were very similar.

6.3.2.2 Fixed Distance Dual KDEs of the Artificial Case and Population Data

The results of the bandwidth and dataset comparisons for the fixed distance dual KDEs performed on the case and population data in *CrimeStat* are shown in **Figure 6.13**. As with the single KDEs, a positive correlation was found between the size of the filter radius and the estimated area of the clustering surface. Dual KDEs using bandwidths around 100*m* appeared to accurately predict the case locations and general areas of the 4 clusters (see the mapped results associated with the h_{AV}, h_{BCV2}, h_{LSCV}, and h_{VC} bandwidths in **Figure 6.13**). Conversely, dual KDEs based on bandwidths greater than 100*m* not only captured the entire clustering area (as opposed to just the case locations), but their estimates extended far beyond the cluster boundaries providing over-estimates of the clustering area. For a given bandwidth, the overall area in which case clustering rates were found appeared to be fairly constant between the non-weighted and weighted analyses. However, the actual rates of case clustering were much higher for the non-weighted data than they were for the weighted data.

The results of the bandwidth and dataset comparisons for the *DMAP* analyses are shown in **Figure 6.14**. As with the fixed distance KDEs performed in *CrimeStat*, a positive correlation was found between the size of the filter radius and the estimated area of the clustering surface. Even though the *Visual Calibration Method* had already identified that a 100*m* bandwidth would best detect the clusters for the Artificial dataset, *DMAP* analyses were still run on the artificial dataset using the bandwidths estimated by the other methods (all of which were larger than 100*m*) in order to have a quantitative measure of the over-estimation of the clustering area. As with the other fixed distance KDEs, for a given bandwidth the overall area in which case

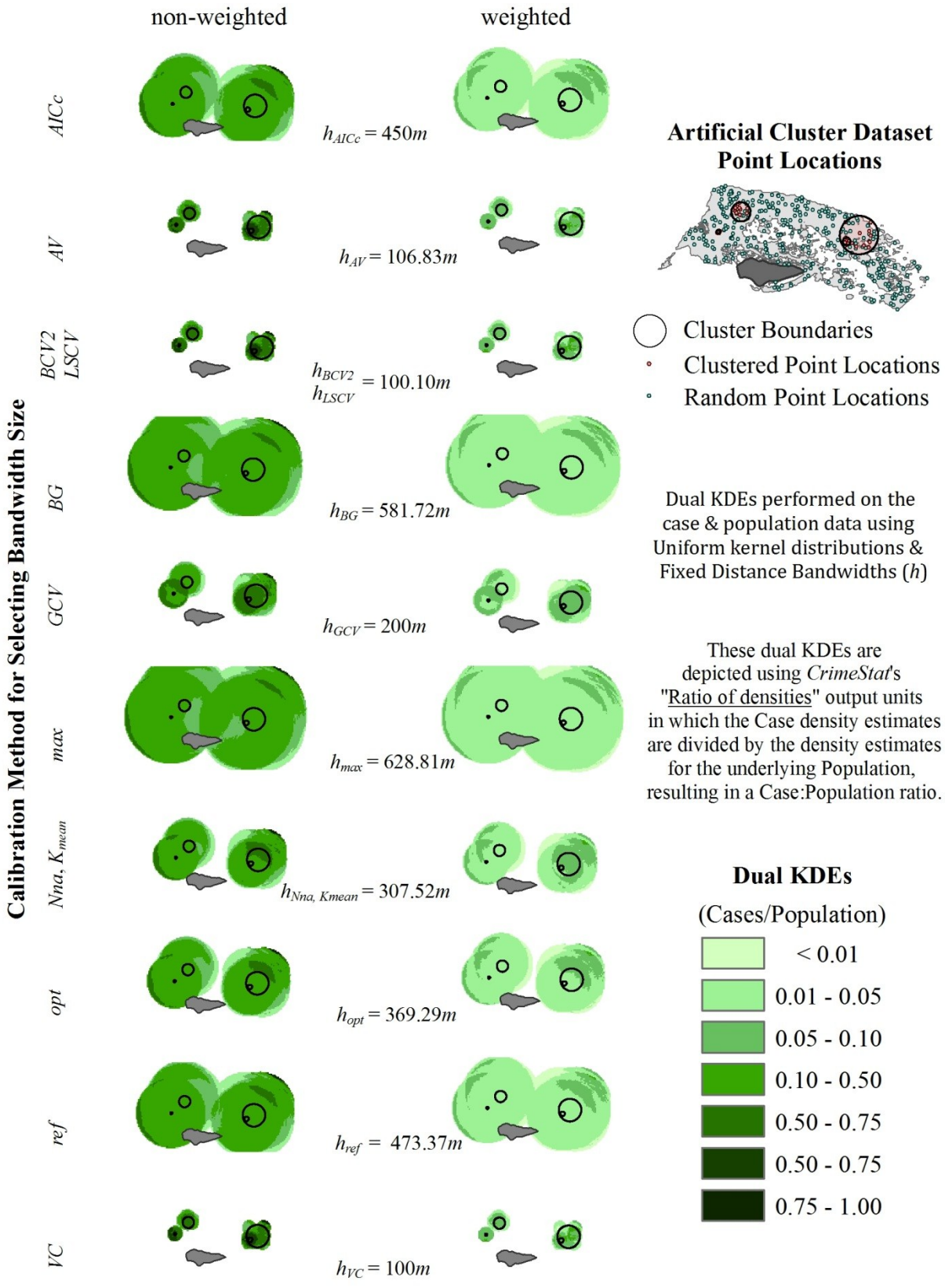


Figure 6.13 *CrimeStat*'s fixed distance dual KDEs of the artificial case and population data.

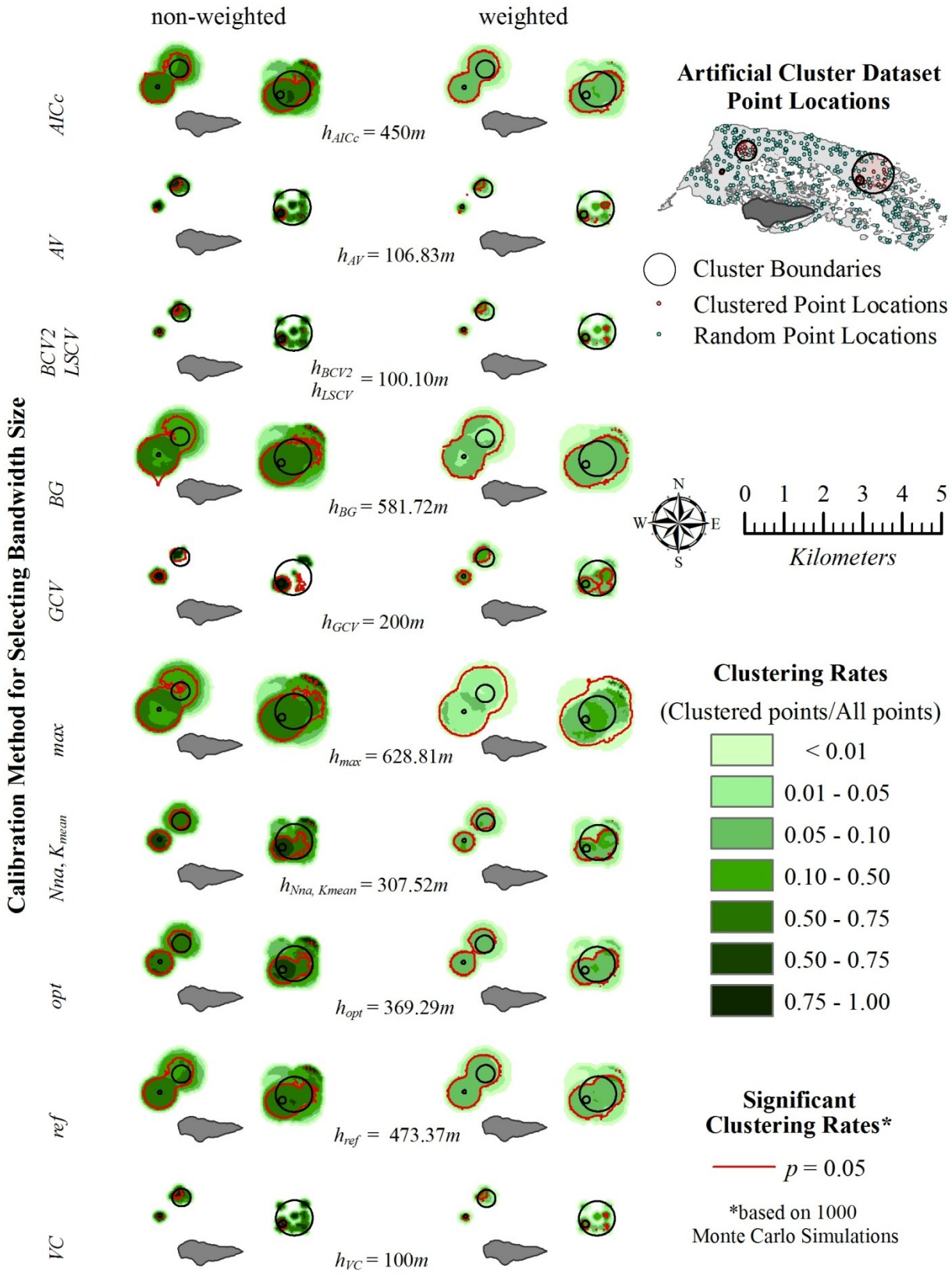


Figure 6.14 *DMAP*'s fixed distance dual KDEs of the artificial case and population data.

clustering rates were found appeared to be fairly constant between the non-weighted and weighted analyses. However, the actual rates of case clustering were much higher for the non-weighted data than they were for the weighted data. When the areas with statistically significant rates of case clustering were compared for a given bandwidth, analyses based on the weighted data had not only larger areas of significant clustering, but also tended to have even lower *p*-values (indicating a greater statistical significance).

DMAP's dual fixed distance KDEs were similar to *CrimeStat*'s dual fixed distance KDEs, in that bandwidths around 100*m* appeared to accurately predict the case locations and general areas of the 4 clusters (see the mapped results associated with the h_{AV}, h_{BCV2}, h_{LSCV}, and h_{VC} bandwidths in **Figures 6.13** and **6.14**). However, the estimated surface area for clustering rates based on bandwidths greater than 100*m* was much smaller for the *DMAP* analyses. As a result, the clustering rates and significant clustering areas based bandwidths less than 500*m* were found to be good predictors of the predefined artificial clustering areas (see the mapped results associated with the h_{AICc}, h_{GCV}, $h_{Nna,Kmean}$, h_{opt}, and h_{ref} bandwidths in **Figure 6.14**). Whereas, dual KDEs based on the same bandwidths in *CrimeStat* had clustering rates that grossly over-estimated the clustering areas. Overall, the h_{opt} and $h_{Nna,Kmean}$ bandwidths appeared to be the "ideal" bandwidths for *DMAP* analysis of both the non-weighted and weighted versions of the artificial dataset. Additionally, while bandwidths ~100*m* were the "ideal" bandwidths for the dual KDEs performed in *CrimeStat*, the *DMAP'*s analyses using the same bandwidths resulted in clustering rates with smaller areas and tended to be more focused on the individual case locations, rather than the four general clustering areas.

6.3.2.3 Adaptive Distance Single and Dual KDEs of the Artificial Data

The results of the single and dual KDEs using kernels with adaptive bandwidths are shown in **Figure 6.15**. To facilitate comparisons between the adaptive bandwidth and fixed

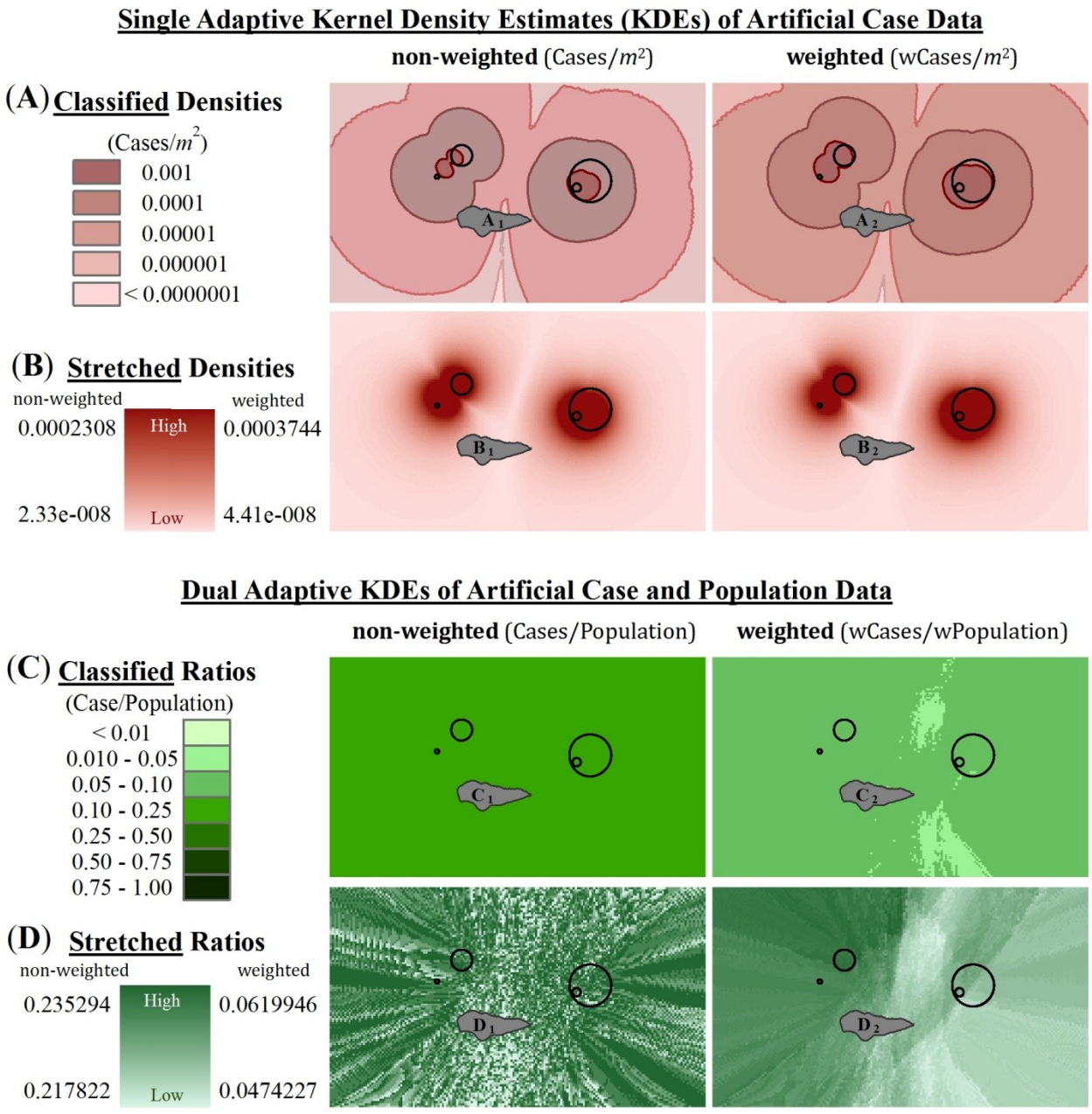


Figure 6.15 Single and Dual Kernel Density Estimates (KDEs) performed in *CrimeStat* using Adaptive distance bandwidths. The Single KDEs were performed on the case data using Quartic kernel distributions and spatially adaptive bandwidths of $h_{Case} \leq 22$ sample points. The dual KDEs were performed on the case and population data using Uniform distribution kernels and spatially adaptive bandwidths of $h_{Population} \leq 187$ sample points. The adaptive distance single KDEs displayed using (**A**) the same classified symbology as the Fixed distance single KDEs shown in **Figure 6.12**; and (**B**) stretched symbology with unique values for the non-weighted ($\mathbf{B_1}$) and weighted ($\mathbf{B_2}$) case densities. The adaptive distance dual KDEs displayed using (**C**) the same classified symbology as the Fixed distance dual KDEs shown in **Figures 6.13** and **6.14**; and (**D**) stretched symbology with unique values for the non-weighted ($\mathbf{D_1}$) and weighted ($\mathbf{D_2}$) Case:Population clustering ratios.

distance bandwidth KDEs all of the single KDEs were performed using quartic kernel distributions, "Relative Density" output units of case density per m^2, and using the same classified symbology in which the density per m^2 ranges from non-zero values < 0.000,000,1 (shown in light red) to 0.001 (shown in a darker red). In this way the results from all of the single KDEs could be directly compared because the only difference between their calculations was their bandwidth. In a similar fashion, all the dual KDEs were performed using Uniform kernel distributions, "Ratio of densities" output units of case densities divided by population densities, and using the same classified symbology in which the Case:Population ratios range from non-zero values < 0.01 (shown in light green) to 1.00 (shown in dark green).

Unlike the single KDEs based on fixed distance bandwidths (**Figure 6.12**), the single KDEs based on adaptive bandwidths of $h_{Case} \leq 22$ points, had interpolated case density values greater than zero for the entire study area (**Figure 6.15A** and **B**). Thus, the estimated area for the case density estimates using adaptive bandwidth kernels was much larger than the estimated case density areas for any of the fixed distance bandwidths, and greatly exceeded the artificial cluster boundaries for both the non-weighted and weighted case data. The general shape of the case clustering areas was circular (**Figure 6.15A**), resembling the output that might be expected from fixed distance KDEs of the same data using a much larger bandwidth than any of the bandwidths tested in **Figure 6.12**. However, despite the extreme over-estimation of the total case clustering areas, the areas with the highest case densities (depicted as the darkest shade of red) for both the non-weighted (**Figure 6.15A$_1$**) and weighted (**Figure 6.15A$_2$**) adaptive kernels did closely approximate the predefined artificial cluster boundaries (depicted by the black circles). In fact, when only the areas with the highest estimated case densities were examined, the adaptive kernel does a better job of detecting the artificial cluster boundaries than the highest and second highest case density estimates of any of the fixed distance bandwidths (**Figure 6.16**).

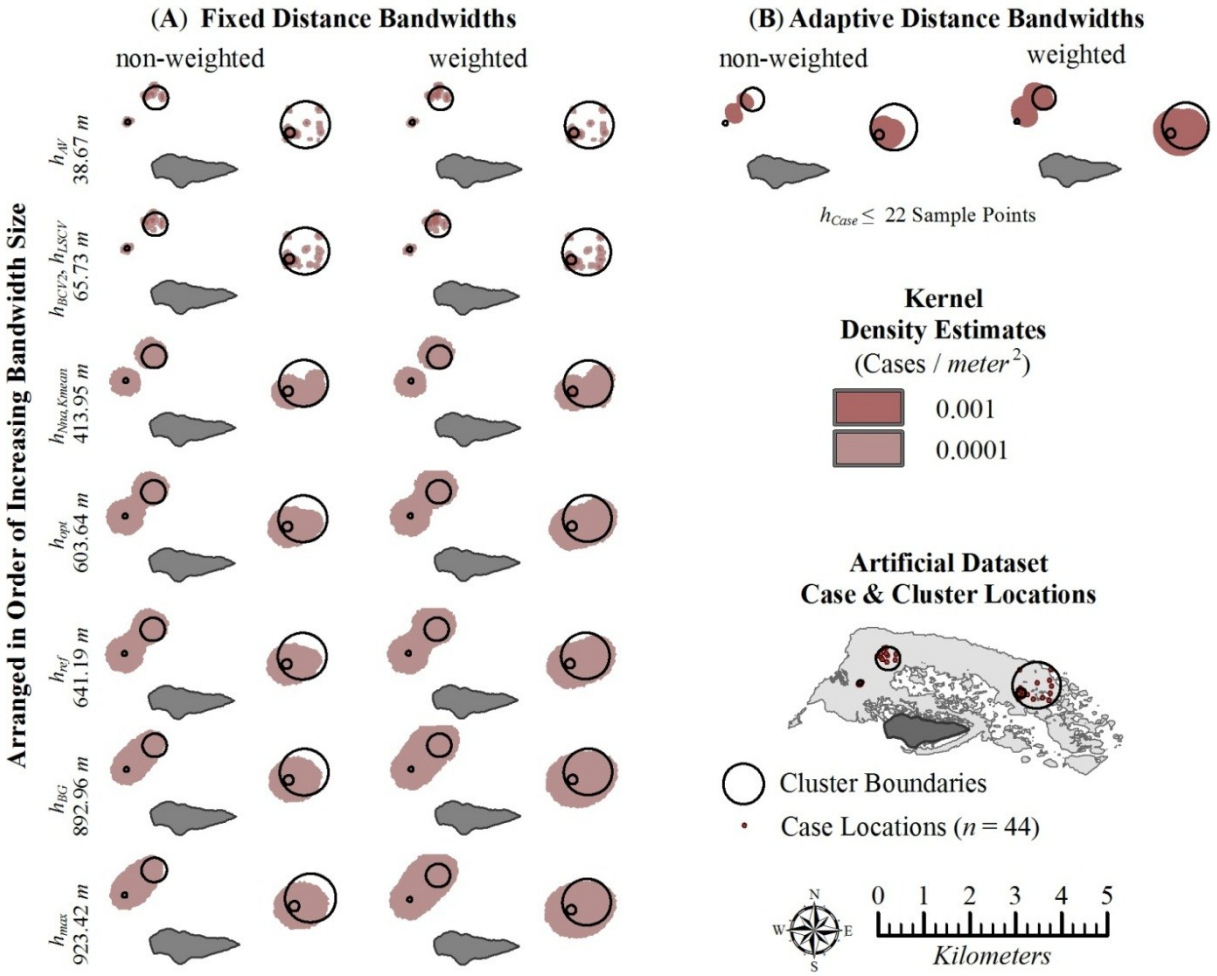


Figure 6.16 Comparing the artificial cluster boundaries to the areas with the highest case density estimates using both fixed (**A**) and (**B**) adaptive distance bandwidths.

As with the adaptive single KDEs, the adaptive dual KDEs based on adaptive bandwidths of $h_{Population} \leq 187$ sampled points, also had interpolated Case:Population density values greater than zero for the entire study area (**Figure 6.15C** and **D**). However, unlike the adaptive single KDEs, the shape of the estimated clustering areas for the adaptive dual KDEs (**Figure 6.15C**) in no way resembled the shapes of the clustering areas for the dual KDEs with fixed distance bandwidths (**Figure 6.14** and **6.15**), much less the locations of the pre-defined artificial clusters. To see whether these results were real, or just an artificial of the symbology used, the single and dual adaptive KDEs were also displayed using "stretched" symbologies based on the minimum and maximum values for the non-weighted and weighted data (**Figure 6.15B** and **D**). While the

stretched symbology did appear to improve the visual accuracy of the single adaptive KDEs (**Figure 6.15B**), it did not improve the dual adaptive KDE, but rather resulted in even stranger depictions of Case:Population clustering ratios (**Figure 6.15D**).

6.3.3 Summary of Spatial Filtering and Smoothing Results

Of all the KDE techniques tested, the single KDEs and the *DMAP* dual kernels with fixed distance bandwidths ≤ 100*m* were the best at detecting the case locations. The pre-defined clustering areas were most accurately predicted by the following kernels: single KDEs performed on the case data using an adaptive bandwidth ≤ 22 points; dual fixed distance KDEs performed on the case and underlying population data using bandwidths ≈100*m* for the *CrimeStat* dual KDEs or between 100 and 500*m* for dual KDEs performed in *DMAP*.

Ultimately, the goal of the spatial filtering and smoothing techniques is to detect areas with higher rates of disease, and when possible, areas that have statistically significant rates of disease. Therefore, it is the locations of the clusters that are of interest, rather than the case locations, given that the locations of the cases are already known. For this reason, when only case data are available, adaptive single KDEs were found to be the most accurate way of detecting the location and size of the artificial clusters. When population data are also available, *DMAP*'s dual KDE technique proved to be far more powerful than *CrimeStat*'s dual KDEs. The case clustering density ratios (or "rates") produced by the *DMAP* analyses appeared to be much more stable than those produced by *CrimeStat,* in that they were less affected by changes in the bandwidth size, and overall all of the *DMAP* rates more accurately reflected the location and size of the pre-defined artificial clusters. In addition, *DMAP* also used Monte Carlo simulations to test the clustering rates in order to determine which areas have statistically significant prevalence rates. Of the 11 bandwidths tested, *DMAP* results based on the $h_{NNA,Kmean}$ (307.52*m*) and h_{opt} (369.29*m*) statistics provided the most accurate cluster detections.

6.4 Spatial Scan Statistics Results

The output from all of the *SaTScan* analyses of the non-weighted and weighted versions of the artificial dataset using the NGO criteria, which is the default criteria for reporting secondary clusters in *SaTScan*, is summarized in **Table 6.3**.

Table 6.3 *SaTScan* results of the non-weighted (**A**) and (**B**) weighted artificial data

(**A**) ***Bernoulli Model***								
Cluster Type	Cluster #	*p*-value	RR	LLR	Observed	Expected	ODE	Radius (km^2)
Primary	1	0.001	11.677419	29.771315	13	1.525333	8.522727	0.166026
Secondary	2	0.001	11.030303	24.913658	11	1.290667	8.522727	0.052679
Secondary	3	0.001	7.479167	15.45491	11	1.877333	5.859375	0.20797
(**B**) ***Poisson Model***								
Cluster Type	Cluster #	*p*-value	RR	LLR	Observed	Expected	ODE	Radius (km^2)
Primary	1	0.084	4.14375	5.64428	9	2.41048	3.733696	0.221536
Secondary	2	0.91	2.903646	1.971386	5	1.80786	2.7657	0
Secondary	3	0.985	2.7556923	1.459074	4	1.50655	2.655072	0.195997
Secondary	4	1	2.704478	0.717677	2	0.753275	2.655072	0
Secondary	5	1	2.248756	0.5011	2	0.90393	2.21256	0.012745
Secondary	6	1	1.41276	0.251212	5	3.615721	1.38285	0
Secondary	7	1	1.21958	0.070059	4	3.31441	1.206851	0.264815

RR=Relative Risk; LLR=Least Likelihood Ratio; Observed= # of observed cases; Expected = # of cases expected; ODE = Observed/Expected

The location, size and significance associated with each of the clusters detected using either *Bernoulli* or *Poisson* probability models and one of four criteria for reporting seconding clusters are visually depicted in **Figure 6.17**. None of the primary or secondary clusters based on analyses performed on the weighted version of the artificial data using the *Poisson* probability model had significant *p*-values associated with them. Conversely, all four of the analyses performed on the non-weighted data using the *Bernoulli model* resulted in three or more significant clusters, which often corresponded with the locations of the pre-defined artificial clusters.

The location and size of the pre-defined artificial clusters were most accurately predicted by the *SaTScan* analysis of the non-weighted artificial data using either the "No Geographical Overlap" (NGO) or the "No Cluster Centers in Other Clusters" (NCCOC) options for reporting secondary clusters. Whereas, the "No Cluster Centers in More Likely Clusters" (NCCMLC) and "No Restrictions" (NR) options resulted in the prediction of far too many clusters, and greatly exceeded the general area of the pre-defined artificial clusters. Overall the results based on either the NGO or NCCOC criteria were found to more accurately represent the artificially clustered data; resulting in identical output for the *Bernoulli model*'s analysis of the non-weighted data, and similar output for the *Poisson model*'s analysis of the weighted data.

6.5 Spatial Autocorrelation (SA) Results

All of the global and local spatial autocorrelation (SA) analyses were performed on weighted versions of the artificial data, in which each artificial location was "weighted" by the number of artificial events present at the given location. SA analyses were not performed on the non-weighted versions of the data; because both the *Getis-Ord G* and *Moran's I* statistics require to be some degree of variation in the attribute values being tested (Getis and Ord 1992; Anselin et al. 2006a; ESRI 2009e,c; Getis 2010). Analyses involving attribute values that are all equal to either one or zero (which is usually the case for presence/absence data) will either render the SA statistic unsolvable or produce unreliable results (ESRI 2009h).

Additionally, both *Getis-Ord G* and *Moran's I* statistics are designed to test associations among "neighboring" features, which means that each feature should share a boundary with each of its neighbors. Following Anselin et al.'s (2006a) example, the artificial point locations were transformed into a continuous polygonal surface through the creation of Thiessen Polygons, in which each polygon represented one point location (see **Figure 5.1** on page 95). The polygons were then "weighted" by assigning the number of individual artificial events associated with

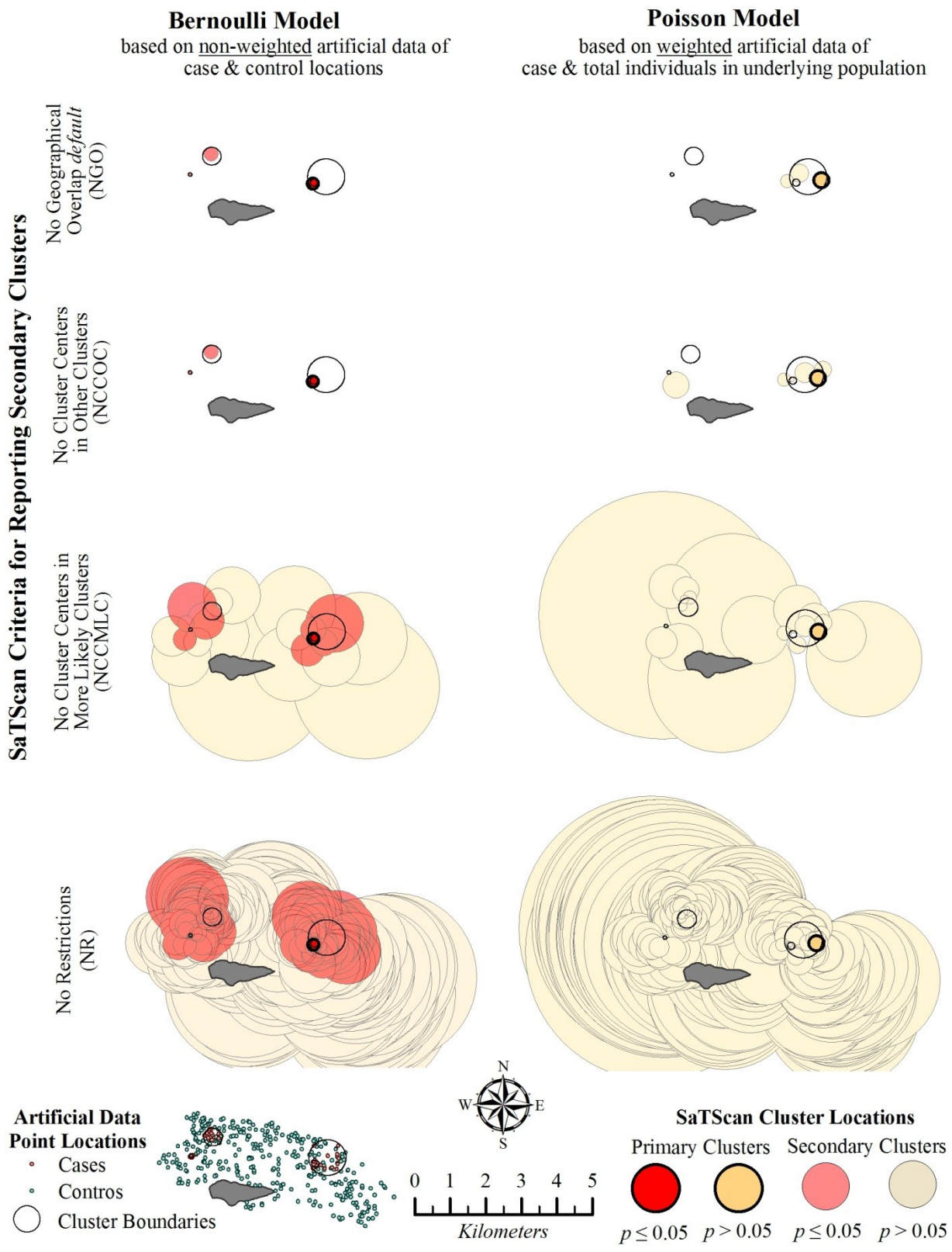


Figure 6.17 Results of all of the *SaTScan* analyses performed on the artificial data.

each location as the attribute value for their respective polygon. Last, in order to insure that the distance calculations used by the SA statistics were calculated correctly, the coordinate system of the artificial polygons had to be transformed from its current Geographic coordinate system (characterized by angular measurements and often depicted in decimal degrees) to a projected coordinate system (characterized by linear measurements and often depicted in metric coordinates). The resulting "input data" for the SA analyses is shown in **Figure 6.18** (**A_1**, **B_1**, and **C_1**).

Based on visual assessment of the spatial distribution of the density (i.e. the mapped weighted attribute values) the following predictions were made. The detection of feature clustering was anticipated for the local SA analyses of both the case and prevalence datasets because the attribute locations for both were limited to the general locations of the artificially defined clusters (depicted by the black circles shown in **Figure 6.18A**, **C**, and **D**). However, feature clustering was not expected for the global SA analysis of either dataset because the artificial clusters were spread across the study area making it unlikely that spatial aggregation would be detected as the region-wide trend. In addition, predominantly low attribute values were associated with both the case (**Figure 6.18A_1**) and prevalence (**Figure 6.18C_1**) features, suggesting any clustering that was detected by local or global analysis would be of low values ("cold spots").

Conversely, little to no spatial association was anticipated for local or global SA analysis of the control data because when the artificial dataset was created, the "Generate Random Points" tool in the *Hawth's Tools* extension was used to insure that the spatial distribution of the control points would be random. The attribute values associated with the control points had also been randomly assigned to each location, suggesting spatial clustering of these values would be unlikely (or "by chance, only"). Visual inspection of the spatial distribution of control density

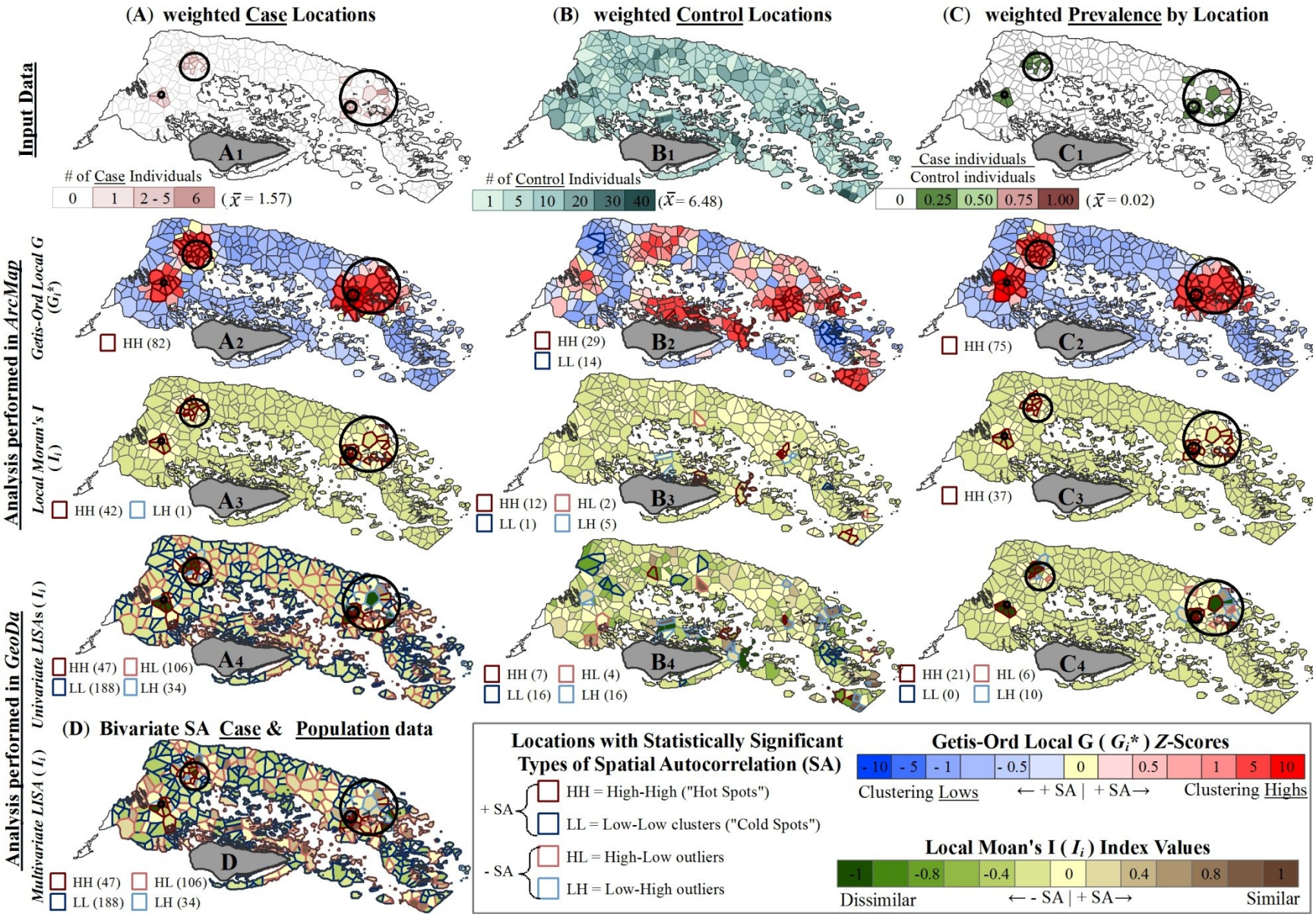


Figure 6.18 Results of the Local Spatial Autocorrelation (SA) analyses performed on the weighted artificial data.

(**Figure 6.18B$_1$**) supported this prediction of zero or negative SA, as the spatial distribution of the density values appeared to be random or possibly even dispersed. Any clustering detected was also likely to be of low values because the mean number of individual control events per feature location was 6.68.

However, the results of both the global and local SA analyses were quite different from what was expected based on the above visual assessment of the spatial distribution of the weighted artificial data. Statistically significant positive SA (clustering) was detected by both the global (**Table 6.4**) and local (**Figure 6.18**) SA statistics for each of the tested datasets. The positive SA detected by the *Getis-Ord General G* (**Table 6.4A**) was of features with high attribute values, indicating that the region-wide trend for each of the artificial datasets was that of High-High (HH) significant clustering. When the *z*-scores and *p*-values for each were compared, the Case and Prevalence features had much higher *z*-scores (25.02 and 27.59, respectively) and lower p-values ($p < 0.001$) than the control ($z = 1.67$, $p = 0.096$) and population ($z = 3.04$, $p = 0.002$) data. This indicates that the clustering of features with high attribute values was much more intense for both the case and prevalence data. The *Global Moran's I* also detected significant positive SA, with substantially different *z*-scores for the two groups (case and presence vs. control and population). The type of clustering (High-High or Low-Low) could not be determined using the *Global Moran's I* because unlike the *General G*, the global version of *Moran's I* does not differentiate between the clustering of high and low values. Instead the *Global Moran's I* is used to determine whether features are more clustered or dispersed.

The similarity between the case and prevalence global *G* results made sense considering that prevalence was calculated as the number of individual case events present at a given location divided by the total number of artificial events present at the same location. As a result prevalence values will only be returned for features case individuals present because

Table 6.4 Results of Global Spatial Autocorrelation analysis of weighted Artificial data

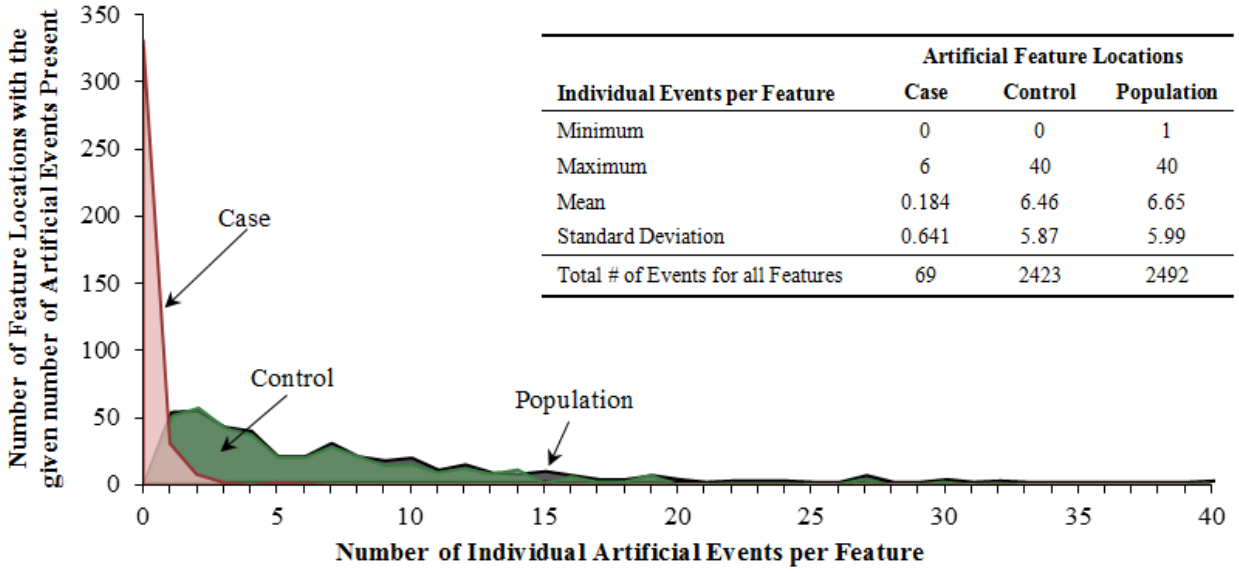
(**A**) ***Getis-Ord General G***

Type of Individuals	Observed General *G*	Expected General *G*	Variance	Significance (*p*-value)	*z*-Score	Description of Spatial Autocorrelation
Case Individuals	0.00183	0.00017	< 0.0000	< 0.001	25.02	Locations with High Case densities are spatially Clustered. There is < 1% likelihood that this clustering is due to chance.
Control Individuals	0.00019	0.00017	< 0.0000	0.096	1.67	Locations with High Control densities are spatially Clustered but not to a significant extent. There is a 5-10% likelihood that this clustering is due to chance alone.
Population (Case + Control)	0.00020	0.00017	< 0.0000	0.002	3.04	Locations with High Population densities are spatially Clustered. There is < 1% likelihood that this clustering is due to chance.
Prevalence (Case/Population)	0.00224	0.00017	< 0.0000	< 0.001	27.59	Locations with High Prevalence rates are spatially Clustered. There is < 1% likelihood that this clustering is due to chance.

(**B**) ***Global Moran's I***

Type of Individuals	Observed Moran's *I*	Expected Moran's *I*	Variance	Significance (*p*-value)	*z*-Score (SD)	Description of Spatial Autocorrelation
Case Individuals	0.59993	-0.00267	0.0006	< 0.001	24.49	Case density locations are spatially Clustered. There is < 1% likelihood that this clustered pattern is due to chance alone.
Control Individuals	0.06914	-0.00267	0.0007	0.005	2.80	Control density locations are spatially Clustered. There is < 1% likelihood that this clustered pattern is due to chance alone.
Population (Case + Control)	0.10956	-0.00267	0.0007	< 0.001	4.37	Population density locations are spatially Clustered. There is < 1% likelihood that this clustered pattern is due to chance alone.
Prevalence (Case/Population)	0.64683	-0.00267	0.0006	< 0.001	26.88	Prevalence rate locations are spatially Clustered. There is < 1% likelihood that this clustered pattern is due to chance alone

$\frac{0\ case\ individuals}{X\ total\ individuals}$ will always result in a prevalence rate of 0. The similarity between the control and population global *G* results was also expected considering control individuals were present at all 375 of the artificial locations and represented ~97% of the underlying population (**Figure 6.19**). The substantial difference between the *Z*-scores of the case (or prevalence) and control (or population) data also made sense considering that the total number of case individuals represented less than 3% of the underlying population. However, the HH clustering detected by the *General G* was not expected; and, after examining the frequency distributions associated with the input data (**Figure 6.19**), did not appear to be justified. The frequency distributions of the



	Artificial Feature Locations		
Individual Events per Feature	**Case**	**Control**	**Population**
Minimum	0	0	1
Maximum	6	40	40
Mean	0.184	6.46	6.65
Standard Deviation	0.641	5.87	5.99
Total # of Events for all Features	69	2423	2492

Figure 6.19 Frequency distributions and descriptive statistics (inset Table) associated with the artificial case, control, and population data.

case data, and to a lesser extent, the control and population data, were all positively skewed indicating that the majority of the artificial feature locations were characterized by low attribute values. The presence of HH clustering might be justified for individual feature locations (detectable through local SA analyses), but in no way seemed to represent the "region-wide trend" for any of the artificial data.

Unlike their global counterparts, the results of the local SA analyses, specifically those analyses performed using *ArcMap,* did appear justified. Local SA analyses of the Case and Prevalence data only detected HH clustering in areas where case individuals were present (**Figure 6.18A$_{2,3}$** and **C$_{2,3}$**). The locations without case individuals present (for both the case and prevalence data) were characterized by the *Getis-Ord Local G* (G_i^*) statistic as having non-significant clustering of low attribute values (**Figure 6.18A$_2$** and **C$_2$**). The *Local Moran's I* (I_i) statistic characterized these same locations as having non-significant negative SA (**Figure 6.18A$_3$** and **C$_3$**). While both the G_i^* and I_i statistics detected statistically significant High-High (HH) and Low-Low (LL) clustering, their analyses resulted in different HH and/or LL estimates

for each of the datasets. The I_i statistic appeared to be much more conservative, detecting roughly half as many HH or LL clusters as the G_i^* statistic. The case and prevalence results were also quite similar for both the G_i^* and I_i statistics. Interestingly, the results of the *local Moran's I* analysis of the same case data differed substantially based on which program it was calculated in (*ArcMap* vs. *OpenGeoDa*). The results of the I_i analyses of the control and prevalence data also differed between programs (**Figure 6.18B$_{3,4}$** and **C$_{3\text{-}4}$**, respectively), but not nearly as much as the case results (**Figure 6.18A$_{3,4}$**). *OpenGeoDa*'s I_i analysis of the case data resulted in statistically significant positive or negative SA for all 375 feature locations. *OpenGeoDa's Bivariate Local Moran's I* (referred to as "Multivariate LISA" in the software) analysis of the case and population data, resulted in the exact same HH, LL, LH, and HL, predictions as *OpenGeoDa's Univariate* analysis of the case data. However, the index values were slightly different for the *Univariate* and *Multivariate* results.

Overall, the spatial distributions of the weighted case, control, and prevalence data were best represented using the local SA available in *ArcMap*. The *Getis-Ord Local G* (G_i^*) and *Local Moran's I* (I_i) provided different information about the nature of the spatial autocorrelation present in each of the artificial datasets. The spatial structure of the datasets were therefore best explained using the results of both the G_i^* and I_i statistics, rather than choosing one over the other. With regard to the *Local Moran's I* statistic, *ArcMap*'s I_i results appeared to be more reliable than *OpenGeoDa*'s Univariate I_i results. There appeared to be no advantage in performing *Bivariate* I_i analysis, as the results produced by this analysis in *OpenGeoDa* were suspect and appeared to be inaccurate. Last, no justification could be found for performing global SA analysis in lieu of local SA analyses, as the summary nature of the global statistics resulted in generalities that did not appear to be representative of the data on which they were based.

Chapter 7. Developing Geospatial Analytical Protocols for Coral Epizootiology Based on the ESDA Results

"Quantifying the spatio-temporal changes in coral reef benthic communities at regional and decadal scales can lead to a broader understanding of the patterns and causes of reef degradation and provide information which will result in more effective management actions (Côté et al. 2005)."
– Schutte et al. (2010)

7.1 Three Tiered Approach to Geospatial Coral Epizootiology

The purpose of spatial epidemiology is to improve scientific understanding of a given disease by incorporating exploratory geospatial analytical methods into a traditional epidemiological framework. As discussed in Chapter 5, a typical spatial epidemiological investigation uses exploratory analytical methods to identify spatial patterns and potential associations between different attributes in the dataset. There are numerous types of exploratory analytical methods which generally fall into one of the following six categories of Exploratory Spatial Data Analysis (ESDA): (1) Mapping and Visualizing data; (2) Point Pattern Analysis; (3) Spatial Filtering and Smoothing; (4) Spatial Scan Statistics; (5) Spatial Autocorrelation; and (6) Spatial Regression. While it is usually a good idea to perform more than one type of analysis on a given dataset, it is often not practical, or appropriate, to perform numerous types of analysis from each of the above ESDA categories.

As mentioned earlier, one of the objectives of this dissertation was to develop geospatial analytical protocols designed to be used by researchers with little to no background in GIS and/or spatial analysis. In order to do this, I began by condensing the six ESDA categories into a simplified and more straightforward 3-tiered approach[1] (see **Table 7.1**). I designed the tiers so that the types of spatial analysis included in each tier all share the same underlying goal, and the complexity and amount of data needed increases with each tier. The first tier contains spatial

[1] The concept of this 3-tiered approach was adapted from Rezaeian et al. (2007) three predominant branches of geospatial epidemiology methods: disease mapping, disease clustering, and ecological analysis.

Table 7.1 Three Tiered Approach to Geospatial Coral Epizootiology

3 Tiers of Geospatial Coral Epizootiology	Spatial Analysis Types according to ESDA Category	Description of the Types of Spatial Information Attained
(1) **Disease Mapping & Visualization**	**ESDA 1. Mapping & Visualizing Data** • Mapping Point Locations using points & polygons • Scaling Point Symbols &/or colors to visualize intensity	**Visualizing Spatial Distributions** • Visualizing the spatial distribution of data locations • Visualizing the spatial distribution of data density (or intensity of an attribute)
(**2**) **Detection & Analysis of Disease Clusters**		
(2_A) **General Disease Clustering** Global spatial statistics assume the spatial distribution of the data are homogeneous & results generally have no spatial output	**ESDA 2. Point Pattern Analysis** **2.1 Centrographic Statistics** • "Mean Center" estimates • *Median Center* (*MdnCntr*) • *Minimum Convex Polygons* (*MCP*) • Standard Distance & Deviation Estimates	**Describe the General Spatial Distribution of the Data** **Shows the Location & Spatial Distribution of Point Patterns** • Identifying the central focal point of the points • Useful when outliers are influencing the mean center • Simplest method for estimating the Home Range of an animal • Estimate the general distribution of the data around a central focal point
	2.2 Distance Statistics • *Nearest Neighbor Analysis* (*Nna*) • *Ripley's K* (*K*)	**Test hypotheses regarding the spatial distribution of points** • Examine spatial dependence (clustering or dispersion) at a given scale • How spatial dependence changes with distance & scales of measurement
	ESDA 5. Spatial Autocorrelation (SA) **5.1 Global SA Analyses** • *Getis-Ord General G* • *Moran's I*	**Whether or Not Clustering is Present** **Whether or not Spatial Autocorrelation (SA) is present region-wide** • Measures the degree of clustering for either "high" or "low" values • Measures the amount of SA based on feature locations & attribute values
(2_B) **Specific Disease Clustering** Local spatial statistics assume the spatial distribution of the data are heterogeneous & there is generally spatial (mappable) output associated with the results.	**ESDA 3. Spatial Filtering & Smoothing** • Single Kernel Density Estimates (KDEs) • Dual KDEs • Dual KDEs with Monte Carlo Simulations	**The Presence, Degree, & Location of Clusters** • Density, Intensity, and Probability estimates • Prevalence, Odds Ratios, & Relative Risk Estimates • All of the above plus Significant Clustering Areas
	ESDA 4. Scan Statistics • Spatial Scan Statistics	**Used to Detect Outbreaks through Spatial Cluster Analysis** • changes in Cluster Size, Significance, & Relative Risk (RR) in a given area
	ESDA 5. Spatial Autocorrelation (SA) **5.2 Local SA Analyses** • *Getis-Ord Local G* (G_i^*) • *Local Moran's I* (I_i)	**Whether or Not Clustering is Present** **Whether or not SA is present, & if so Where is it Occurring** • Identifies where "high" or "low" values cluster spatially • Identifies the locations of high & low clusters, as well as spatial outliers
(**3**) **Disease Modeling, Prediction, & Ecological Analysis**	**ESDA 4. Scan Statistics** • Space-Time Scan Statistics • Temporal Scan Statistics	**Used to Detect Outbreaks through Temporal Cluster Analysis** • changes in Cluster Size, Significance, & RR in a given area over time • changes in Cluster Size, Significance, & RR in over a specified time period
	ESDA 6. Spatial Regression Analyses (RA) • *Ordinary Least Squares* (*OLS*) Regression • *Geographically Weighted Regression* (*GWR*)	**Performs Local RA without assuming Spatial Homogeneity** • OLS results output are used to build the GWR model • Assesses spatial heterogeneity between independent & dependent variables

methods and visualization techniques used to visualize the spatial distribution of diseased corals through the creation of different types of maps. The second tier contains spatial methods that are designed to detect and analyze spatial clusters of diseased individuals. Last, the third tier contains spatial methods used to model the relationship between the spatial distribution of diseased corals and other spatial, temporal, and ecological variables, in order to better understand how these variables influence the spatial nature of a given coral disease, test various hypotheses, and possibly even predict future disease outbreaks. The following sections provide a brief overview of the types of spatial analysis included in all three tiers that I recommend using based on the results of the ESDA techniques used on my artificial cluster dataset, which were described in depth in the previous chapter (Chapter 6).

7.1.1 Tier 1: Disease Mapping and Visualization

Maps are an important first step in any spatial epidemiological investigation because they allow the spatial distribution of phenomena (such as the locations of disease corals) to be visually observed (Câmara et al. 2008). In some cases, showing *where* an event is occurring may also provide insight as to *why* the event is occurring (Waller and Gotway 2004). It is important to note that the visualization options for both the map and the representation of the data locations within the map should be based on the nature of the data, the type of study, and the overall purpose of the map (i.e. what message is the map intended to convey).

Maps of non-weighted data locations are convenient in that both case and control locations can be displayed in the same map (**Figure 7.1A**). However, they can be misleading because they imply that all the individuals present at case locations are case individuals, which is often not the case. Maps in which the locations are weighted by the number of individuals present provide a more realistic depiction of the underlying densities. Unfortunately, case and control densities cannot be displayed on the same map without overlapping values being

obscured (**Figure 7.1C**). Prevalence maps offer a way of visualizing case density while also accounting for the underlying population density in the same map (**Figure 7.1B**). The drawback of prevalence maps is that, all information regarding the population densities of non-case locations is lost. Following this logic, point-based visualization techniques are recommended for either depicting feature locations on their own, or when dealing with non-weighted data (**Figure 7.1A**). When weighted datasets are available, separate polygon-based mapped depictions of case prevalence (**Figure 7.1B**), and case and control density should be used (**Figure 7.1C**).

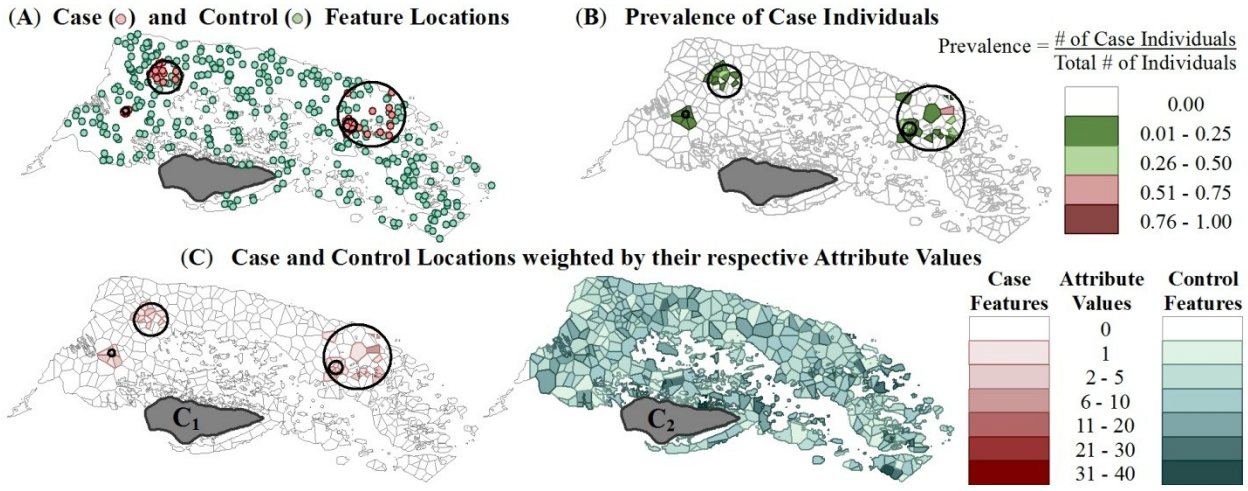


Figure 7.1 (**A**) non-weighted Case and Control feature locations displayed in the same map. (**B**) Prevalence Map, weighting each feature by case prevalence estimates. (**C**) Case ($\mathbf{C_1}$) and Control ($\mathbf{C_2}$) features locations weighted by their respective attribute values.

7.1.2 Tier 2: Detection and Analysis of Disease Clusters

The detection and analysis of spatially clustered diseased corals is a crucial component of most spatial epidemiological investigations because these disease "hotspots" can then be further examined using more traditional epidemiological methods to investigate the epidemiology, etiology, pathology, and pathognomy of the disease (Berke 2004). This "hotspot-to-causation" approach enables a more comprehensive and less subjective way for epidemiologists to examine the spatial aspects of both the disease transmission and infection rates (Chaput et al. 2002).

There are two types of cluster detection, which divides this tier into the following two sub-tiers: (2_A) General Disease Clustering; and (2_B) Specific Disease Clustering. General disease clustering methods use global statistics to detect and analyze the "overall clustering tendency of the disease incidence in a study region" (Rezaeian et al. 2007). Whereas, specific disease clustering methods use local statistics to detect and analyze the locations of *specific* disease clusters within the study region.

7.1.2.1 Tier 2_A: General Disease Clustering

The following ESDA categories contain types of spatial analysis that are considered to be global statistics: ESDA Category 2, Point Pattern Analysis (PPA); and ESDA Category 5, Spatial Autocorrelation (SA). Both of the PPA sub-categories, (centrographic statistics and distance-based statistics) and one of the SA sub-categories (global SA) are considered to be measures of general disease clustering because the types of spatial analysis included within each of them are used to search for region-wide trends. Of the eight types of analysis included among these three sub-categories (see **Table 7.1**), only the *Ripley's K* statistic was considered to be robust enough to recommend using on coral health data.

The *Ripley's K* Statistic

The *Ripley's K* statistic is a global measure designed to determine the spatial scale at which clustering is present on the landscape, but it does not identify where on the landscape the clustering is occurring (Gatrell et al. 1996; Lancaster and Downes 2004; Marcon and Puech 2009; Bayard and Elphick 2010). In many types of spatial analysis, the scale or resolution of the data tends to have a strong influence on the appearance of the resulting analyses. The *Ripley's K* statistic can be used to study how the spatial dependence present within a given set of points changes across multiple distances (Bailey and Gatrell 1995; Lentz et al. 2011).

Unlike, the other PPA statistics, the spatial distribution of the underlying population can also be taken into account when studying case data, by using the difference function (*D*) to subtract the *Ripley's K* population estimates from the *Ripley's K* case estimates. The *Ripley's K* results provided new information about how the different distance scales affected the estimated spatial dependence (clustering or dispersion) of both the non-weighted and weighted artificial case and population data. Additionally, numerous hypotheses regarding the spatial distribution of both the artificial case and population data were able to be tested through the use of the difference function or by simply plotting different combinations of the *Observed K* and confidence interval results.

The *Ripley's K* statistic has been used in previous coral disease studies to facilitate a better understanding of the etiologies of their respective diseases by examining the spatial disease distribution, and testing hypotheses regarding the mode of transmission and infection (Jolles et al. 2002; Foley et al. 2005; Zvuloni et al. 2009). The major drawback of the *Ripley's K* statistic is that it provides no spatial output of *where* the clustering is occurring. For this reason, I recommend that the *Ripley's K* statistic only be used in conjunction with other ESDA techniques that provide spatial output (such as those included in the following sub-tier).

7.1.2.2 Tier 2_B: Specific Disease Clustering

The following ESDA categories contain types of spatial analysis that are considered to be local statistics: ESDA Category 3, Spatial Filtering & Smoothing; ESDA Category 4, Scan Statistics; and ESDA Category 5, Spatial Autocorrelation (SA). All of the types of Kernel Density Estimates (KDEs) included under the Spatial Filtering & Smoothing ESDA category are considered Specific Disease Clustering methods, given that they all produce mappable results that provide spatial estimates of disease surfaces. Of the three types of Scan Statistics in ESDA Category 4 (Spatial, Temporal, and Spatio-Temporal), only the Spatial Scan Statistical methods

were included in this tier; the temporal nature of the two other scan statistics makes them better suited for Tier 3's disease modeling and predictive analyses. Last the types of analysis included under the second SA sub-category (local SA) was considered to be measures of specific disease clustering because both types of spatial analysis included within this sub-category are used to locate disease clusters. Of the six types of analysis included among these ESDA categories (see **Table 7.1**), only the Dual KDEs performed in *CrimeStat* were excluded from the recommended methods of detecting the locations of disease clusters in coral health data.

Spatial Kernel Density Estimates (KDE)

Spatial kernel density estimates (KDEs) are often computed using a process known as "spatial filtering," which is a type of non-parametric, graphical analysis used to calculate the predicted value at a given point based on the values of the surrounding data points (Carlos et al. 2010; Cai et al. 2011). These spatial filters use probability distribution functions (known as "kernels") to smooth some of the variability and noise in the dataset without losing the local features of the data, resulting in the creation of smooth, continuous maps of density estimates (Williamson et al. 1998; Waller and Gotway 2004; Anselin et al. 2006b; Carlos et al. 2010; Cai et al. 2011).

Spatial KDEs require the user to define the size of two spatial parameters: the size of the filter radius (also referred to as "bandwidth"), and a grid of the study area with a defined cell size. Spatial parameter selection is a critical first step as the finest bandwidths (for example the immediate area around one coral colony) will lack the potential to identify local areas of disease clustering. Whereas, parameters that are too coarse will overestimate the disease surface and often under estimate the severity of localized clusters (Hall and Marron 1991; Hazelton 1996; Jones et al. 1996; Danese et al. 2008). The size of the grid cell is also important because it is what enables identification of the clustering patterns, if the grid cell is too small the interpolation

will become jagged, while an excessively large grid cell will lose the fine-scale detail (Rodgers and Carr 2002; Wiegand and Moloney 2004; Chainey and Ratcliffe 2005; Danese et al. 2008; Ratcliffe 2010).

KDEs can be performed as a univariate analysis of a set of point locations, or as a multivariate analysis of 2 sets of point locations. The former type of analysis is referred to as a "single" KDE and is most commonly performed on data regarding the locations of case events; while, the latter is referred to as a "dual" KDE & can be used on various types of "numerator"& "denominator" data. "From an epidemiological perspective, kernel estimation is of most value in estimating the intensity of one type of event relative to another" (Carlos et al. 2010); such as, comparing the spatial densities of case locations to that of the underlying population at risk. This makes it possible to differentiate "real" case (disease) clusters and areas at greater risk of becoming diseased. Conversely, single KDEs of just case data run the risk of identifying false clusters resulting from underlying population dynamics (Bithell 1990; Rushton and Lolonis 1996; Levine and Associates 2004; Waller and Gotway 2004; Anselin et al. 2006b; Smith and Bruce 2008; Carlos et al. 2010).

When there is only data on the diseased corals, I recommend using *CrimeStat*'s single adaptive-distance KDEs with an adaptive bandwidth equal to no more than half the total number of disease locations (see **Figure 7.2A**). However, when there is data on both the diseased corals and the underlying coral population at risk, I recommend using the *DMAP*'s fixed-distance dual KDEs, using an Optimized Bandwidth (calculated using the h_{opt} statistic on the locations of the population data), and 1,000 Monte Carlo simulations (see **Figure 7.2B**) . In both cases the grid cell resolution should be slightly larger than the size of each survived data point. For example, the points in the artificial dataset represent 10x30*m* transects, and the optimum grid cell size was found to be $50m^2$.

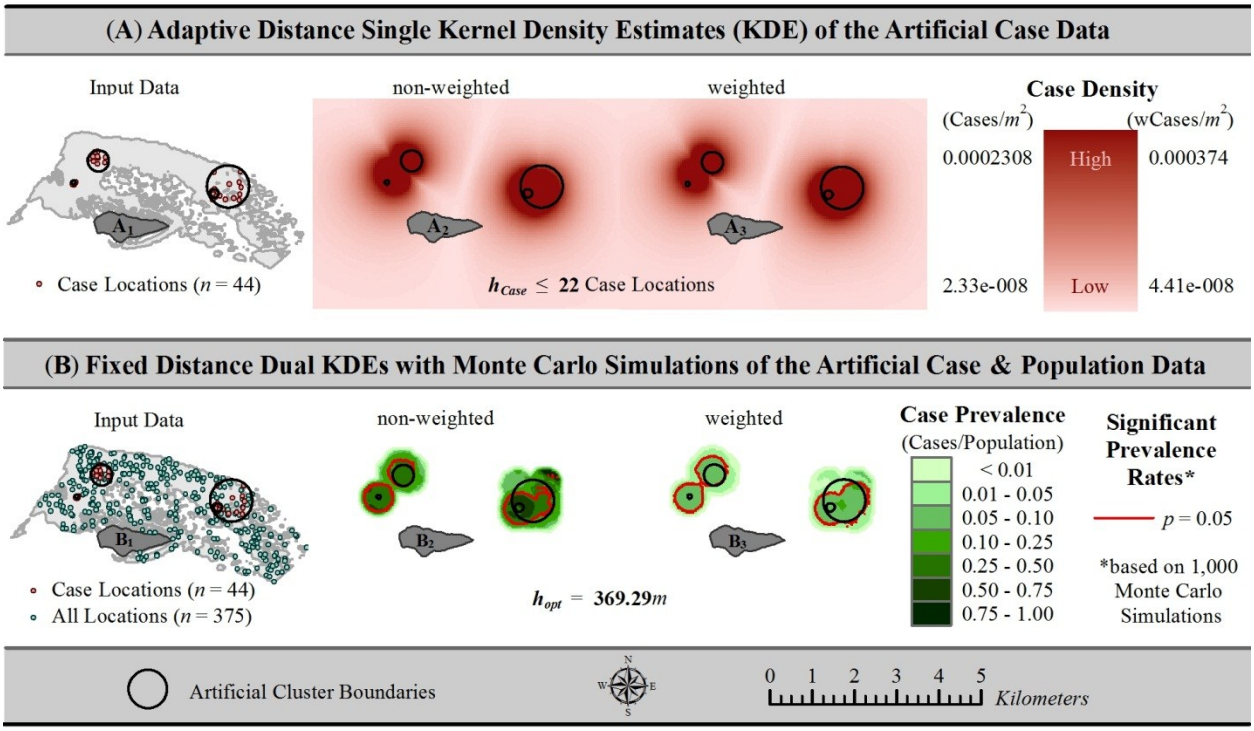


Figure 7.2 Recommended spatial filtering and smoothing methods based on data availability. (**A**) When only case data are available, adaptive-distance single Kernel Density Estimation (KDE) should be performed in *CrimeStat* using Quartic kernel distribution, a spatially adaptive bandwidth greater than or equal to half the total number of case locations, and then displayed in *ArcMap* using the Stretched density symbology setting. (**B**) When both case and population data are available, fixed-distance dual KDEs should be performed in *DMAP*, using 1000 Monte Carlo Simulations. The Optimized Bandwidth (h_{opt}) estimation method should be used in conjunction with the population location data to calculate the appropriate "filter radius" for both the numerator (case) and denominator (population) data in *DMAP*. The clustering rates (case prevalence) should then be displayed in *ArcMap* using either the Stretched or Classified (shown above) symbology, with the areas with statistically significant clustering depicted as contours.

Spatial Scan Statistic

Spatial Scan Statistics are one of the most common types of spatial analysis used by spatial epidemiologists to detect the locations of current disease clusters (Kulldorff et al. 2005; Robertson and Nelson 2010). As mentioned previously, there are three main types of scanning statistical analyses provided in *SaTScan*: purely spatial, purely temporal, and space-time analyses (Kulldorff 2010). Purely spatial scanning statistical analysis takes only the locations (and potentially weights) of the numerator (case) and denominator (population) into account, ignoring

any temporal information that may be included in the dataset. Overall the results based on either the No Geographical Overlap (NGO) or No Cluster Centers in Other Clusters (NCCOC) cluster detection criteria were found to more accurately represent the artificially clustered data; resulting in identical output for the *Bernoulli model*'s analysis of the non-weighted data, and similar output for the *Poisson model*'s analysis of the weighted data.

While *SaTScan*'s scanning statistics offer quick results indicating the locations of significant disease clusters, it is important to note that this software can only be used when data are available for both the diseased corals and the underlying population at risk. Additionally, the results only indicate the locations of significant disease clusters, providing no information about any possible spatial variation present in the disease distribution. For this reason *SaTScan*'s purely spatial analysis should either be used in conjunction with the other cluster detection methods as a means of explaining a given disease distribution, or it used at the beginning of a study so that microbial analyses can be performed using *a priori* knowledge of the locations of significant disease clusters. However, it should not be used to replace any of the other cluster detection and analysis methods, or even the methods from Tier 1; as the *SaTScan* analyses alone do not provide enough information on the spatial nature of a given disease outbreak.

Local Spatial Autocorrelation (SA) Analyses

Spatial autocorrelation (SA) is a statistical measure of the degree of spatial association between the geographic locations and attribute values associated with each feature in a given dataset (Goodchild 1988; Fotheringham et al. 2000,2002; de Smith et al. 2009; Griffith 2009; Myint 2010; Fischer and Wang 2011). Spatial patterns in which a feature with a given attribute value are surrounded by features with similar attribute values (such as a location with a high number of diseased coral colonies surrounded by locations that also have high numbers of diseased colonies present) show "positive" spatial autocorrelation.

Conversely, when the attribute values associated with neighboring features tend to be more dissimilar than the values of features located further away (such as a location with a high number of diseased coral colonies surrounded by locations that have little to no diseased colonies present), then the spatial pattern as a whole shows "negative" spatial autocorrelation. Last, when the distribution of the attribute values appear to be independent of the geographic location of their associated features, suggesting the attribute values and feature locations are not spatially autocorrelated, but rather appear to have a random spatial distribution.

Unlike the previous two types of specific disease cluster detection and analysis methods, local spatial autocorrelation (SA) methods require that the point data be transformed into a continuous polygonal surface through the creation of Thiessen Polygons. Additionally, local SA methods require weighted data, as SA analyses involving attribute values that are all equal to either 1 or 0 (which is usually the case with presence/absence data of non-weighted locations) usually results in either unreliable or even unsolvable results (ESRI 2009h). While local SA analyses can be performed using either *ArcMap* or *OpenGeoDa*, the results of the local SA analyses of the artificial cluster dataset found the spatial distributions of the artificial case, control, and case prevalence data were best represented by the local SA analyses performed in *ArcMap*. Additionally, I recommend using both the *Getis-Ord Local G* (G_i^*) and the *Local Moran's I* (I_i), as they each provided different information about the nature of the spatial autocorrelation present in each of the datasets.

7.1.3 Tier 3: Disease Modeling, Prediction, and Ecological Analysis

Disease modeling, prediction, and ecological analyses are grouped together in the third and last tier because they all build on the results from the previous two tiers (primarily the results from Tier 2). For example once the spatial distribution of a given disease has been identified, models can be used to try to explain why the disease is distributed this way, and test hypotheses

about other factors that may be influencing the spatial distribution of the disease. Once these models have been calibrated such that they appear to be accurately representing the current disease distribution, additional temporal information can be added to try to identify both past and future disease outbreaks. While none of the types of analysis that fall under this tier were tested in this dissertation (due to the limited nature of the artificial dataset), recommendations can still be made as to how they should be used with respect to coral health data.

Space-Time and Temporal Scan Statistics

SaTScan's scanning statistics offer quick results indicating the presence of significant disease clusters, however, this software can only be used when data are available for both the diseased corals and the underlying population at risk. As mentioned in the previous section, *SaTScan* should not be used to replace other methods in this Tier, unless there is not enough data available to perform the other analyses. For example, say a researcher has a coral health dataset in which data were collected on both the diseased and underlying population of corals at risk, in a given location over a specified time period (days, weeks, months, years, etc.), but no spatial information was collected on any of these corals. In this situation, out of all the types of analysis discussed in this dissertation *SaTScan*'s Purely Temporal analysis would be the only type of analysis possible. Or if a researcher had a coral health dataset that contained both spatial and temporal information on diseased and the underlying coral population at risk for a given location over a specified period of time, but the dataset was not suitable for spatial regression analyses because it contained less than 300 total data points and/or no ecological data. In this case, given that the dataset does not contain enough feature locations or independent variables to perform a spatial regression analysis or any other type of ecological analysis, then *SaTScan*'s Space-Time scanning statistic would likely be the best type of analysis to model changes in disease clustering over time, and possibly even predict the locations of future disease outbreaks.

Spatial Regression

Spatial regression is often the final step in ESDA because it goes beyond just visualization and cluster detection, allowing the relationships between different spatial variables to be modeled in order to better explain which factors (independent variables) have the most influence on the spatial nature of the dependent variable. Well specified regression models can be used to explain the phenomena of interest, test hypotheses, and potentially even predict future outcomes (ESRI 2009f; Rosenshein et al. 2011).

Spatial regression is designed to evaluate and model the spatial relationship between two or more attributes associated with a minimum of several hundred-feature locations (ESRI 2009d). Regression techniques are used to model the relationship between a dependent variable and one or more independent variables. The dependent variable is the variable or process that is trying to be understood (such as the spatial distribution of diseased corals), and the independent variable(s) are factors (such as other coral diseases, environmental data, anthropogenic stressors, etc.) that might help explain the spatial distribution of the dependent variable (Charlton and Fotheringham 2009). In this way spatial regression models could be used to associate disease clusters with surrounding environmental factors.

Ecologic Analysis

Ecologic Analysis refers to methods which look for associations between disease incidence (i.e. coral disease) and other social (i.e. other coral diseases, or competition for space among corals, etc.), environmental (surface currents, sea surface temperatures, wind direction, salinity, etc), and anthropogenic (human population size, pollution, frequently visited tourist sites, etc.) covariates (Rezaeian et al. 2007). Given this definition, ecologic analyses can be performed during spatial regression.

7.2 Proposed Analytical Protocols for Geospatial Coral Epizootiology

Geospatial methods are both data and situational dependent. For this reason, before performing any type of spatial analysis a conscious, informed decision needs to be made regarding the type of analysis (and the spatial parameters that will be used during the analysis) that is most appropriate for the given study. The following sections provide brief descriptions of the different analytical protocols I have developed using this three tiered approach to geospatial coral epizootiology. These protocols have been designed to be used on existing coral health datasets.

7.2.1 Recommended Protocols When Only Coral Disease Data are Available

At present, many investigations examining spatial data concentrate (and more importantly, collect) only on the variable of interest. In the case of coral disease, this would be the location of the diseased coral. However, without similarly collected denominator data, it is impossible to know if the pattern revealed by the analysis is a disease "hotspot," or simply indicative of locations with higher densities of coral (i.e. the more coral there is, the more diseased corals are likely to be found).

Despite the potential for misleading results, existing datasets containing only coral disease (case) data should not be abandoned all together. **Table 7.2** provides the types of spatial analysis I recommend performing on either non-weighted or weighted case data. In which "non-weighted case data" refers to datasets that contain presence/absence type data providing only the locations in which a given coral disease was found to be present, but with no indication of how many corals were found at this location. Whereas, "weighted case data" refers to datasets that contain information on the number (or percentage) of diseased coral colonies at each location.

In both cases, the researcher should begin with Tier 1, by creating a map to visualize the spatial distribution of the diseased corals. Non-weighted case data are best represented using

point-based maps, in which all the diseased locations are represented using the same cartographic symbol (for an example see how the non-weighted artificial case locations were depicted by filled red dots in **Figure 7.2A$_1$**). Conversely, when weighted case data are available, I recommend using polygon-based maps, in which each polygon represents the location where diseases corals were found to be present, and the number of diseased corals within each location is depicted using different shades of the same color; for an example see how the weighted artificial case locations were depicted using different shades of red in **Figure 7.1C$_1$**.

Next, the researcher should move on to Tier 2's recommended methods for detecting and analyzing disease clusters (listed in **Table 7.2**). First *Ripley's K* analyses should be performed on the non-weighted and (if possible) weighted versions of the coral disease data. The results from these analyses should then be brought into a spreadsheet and graphing program (such as Microsoft's Excel). The *Observed K* [$L(d)$] values should then be transformed into *Normalized K* [$L(d) - d$] values by subtracting the distances (d), also referred to as the "*Expected K*" values; likewise, the upper and lower confidence intervals (*CI*) should also be transformed into *Normalized CI* [$CI - d$]. Datasets which contain only non-weighted coral disease data should then plot the *Normalized Observed K* values [$L(d)_{\text{Cases}} - d$] and *Normalized CI* [$CI_{\text{Cases}} - d$] for their data on the y axis, against the tested distance (d) on the x axis (resulting in a graph similar to the one shown in **Figure 6.6A** on page 136). The resulting plot can then be used to determine the following: (1) whether the locations with disease present appear to be spatially aggregated (clustered) or spatially dispersed (not clustered); (2) does their anticipated spatial distribution change based on what distance it is calculated at; and (3) is any statistically significant aggregation or dispersion present and if so what distances does it occur at?

Table 7.2 Recommended Types of Spatial Analyses when Only Coral Disease (Case) Data are Available.

Geospatial Coral Epizootiology Tiers & ESDA Types	non-Weighted Case Data	Weighted Case Data
(1) Disease Mapping & Visualization[1]		
• Point-based maps (using circular, dot-shaped symbols)	Dots of uniform size & color (ex. filled dots) (see **Figure 7.2A$_1$**)	—
• Polygon-based maps (using Thiessen Polygons)	—	Color-ramp polygon shading based on weight (see **Figure 7.1C$_1$**)
(2) Detection & Analysis of Disease Clusters		
(2$_A$) General Disease Clustering (no spatial output)		
2. Point Pattern Analysis (PPA) Distance Statistics[1]		
• ***Ripley's K*** [$L(d)$]		
Results *Ripley's K* analyses are plotted in excel using normalized *Observed K* values [$L(d) - d$], normalized upper & lower Confidence Intervals [$CI - d$] normalized *Expected K* values of [$d - d$] or $y = 0$	Plot distance (x axis) vs. the following: $L(d)_{Cases} - d$ & $CI_{Cases} - d$ (see **Figure 6.6A**)	Plot distance (x axis) vs. the following: $L(d)_{wCases} - d$ & $CI_{wCases} - d$ (see **Figure 6.6B**) $L(d)_{wCases} - d$ & $CI_{Cases} - d$ (see **Figure 6.8A**) $L(d)_{Cases} - d$ & $CI_{wCases} - d$ (see **Figure 6.9A**)
(2$_B$) Specific Disease Clustering (spatial output)		
3. Spatial Filtering & Smoothing		
○ **Single Kernel Density Estimates** (sKDEs)[2]	sKDEs performed on Case locations using Quartic distribution spatial kernels & Relative Densities output units	
• Adaptive distance bandwidths	select "Adaptive" as the choice of bandwidth & set the minimum sample size equal to half the total # of case locations	
	make sure the "use weighting variable" box is not selected (see **Figure 7.2A$_2$**)	make sure the "use weighting variable" box is selected & that the data column containing the weighted data are selected under "Weight" on the input data screen (see **Figure 7.2A$_3$**)
5. Spatial Autocorrelation (SA)		
○ **Local SA analyses**[1]	—	Thiessien Polygons generated from Case Locations The resulting index values, cluster types, & cluster significance (p-values) for each analysis are mapped (see **Figures 6.18A$_{2\text{-}3}$**)
• ***Getis-Ord Local G*** (G_i^*)		
• ***Local Moran's I*** (I_i)		

Superscript numbers indicate which software to perform the analysis in: 1 = *ArcMap* 2 = *CrimeStat*

Datasets that contain weighted coral disease data should compare the results of the above plot based on the non-weighted version of their data to the same type of plot using the weighted version of their data, in which $L(d)_{\text{wCases}} - d$ and $CI_{\text{wCases}} - d$ are plotted against d (resulting in a graph similar to the one shown in **Figure 6.6B** on page 136). By comparing these two plots, the researcher can see what effect, if any, taking the underlying diseased coral colony data into account has on the anticipated spatial distributions of the given coral disease. This researcher could then go on to test additional hypotheses by plotting the *Normalized Observed K* values of either the weighted versions of their data with the *Normalized CI* values from their non-weighted data, and visa versa. For example, the null hypothesis that the spatial distribution locations weighted by the number of diseased colonies within them would not be significantly different than the underlying spatial distribution of just the locations would be tested by plotting the *Normalized Observed K* values for the weighted data [$L(d)_{\text{wCases}} - d$] along with the *Normalized CIs* for the non-weighted data [$CI_{\text{Cases}} - d$] (see **Figure 6.8A** on page 138). In order for this null hypothesis to be accepted the line representing the *Normalized Observed K* values for the weighted data must fall inside the *Normalized CIs* for the non-weighted data. Or by plotting the opposite combination of the *Normalized Observed K* values for the non-weighted data [$L(d)_{\text{Cases}} - d$] along with the *Normalized CIs* for the weighted data [$CI_{\text{wCases}} - d$] (see **Figure 6.9A** on page 140), the null hypothesis that the spatial distribution of the locations weighted by the number of diseased corals within them would be more clustered or dispersed through chance alone. In this case, the null hypothesis would only be accepted if the *Normalized Observed K* values for the non-weighted data did fall within the confidence intervals based on the weighted data.

Once the researcher has finished the *Ripley's K* analyses, they should go on to use specific disease clustering methods in order to see where exactly these disease clusters are

located. First the researcher should perform a single kernel density estimate (KDE) using an adaptive distance bandwidth based on half the total number of diseased locations. The only difference in the analytical procedure for a researcher who has non-weighted disease data and a researcher who has weighted disease data would be whether or not they select the "use weighting variable" box in *CrimeStat*. The results from this analysis would then be brought into *ArcMap* to be displayed using a stretched symbology, in which the darkest areas represent locations with the highest coral disease densities (see **Figures 7.2A$_2$** and **7.2A$_3$**).

Depending on how satisfied they were with the results of the single KDE, the researcher with weighted coral disease data, also would then have the option of also performing Local Spatial Autocorrelation (SA) analysis on the weighted version of their dataset. If they decide to do this, their disease location data would need to be transformed from point-based data into Thiessien polygons which can be done using the *XTools* extension in *ArcMap*. I then recommend performing both the *Getis-Ord Local G* (G_i^*) and *Local Moran's I* (I_i) analyses, which are referred to in *ArcMap*'s "Mapping Clusters" toolset as "Hot Spot Analysis (Getis-Ord Gi*)" and "Cluster and Outlier Analysis: Anselin's Local Moran's I," respectfully. The results of both of these analyses offer additional information about the spatial nature of the disease clustering, such as whether each location would be characterized as having: Positive SA, and if so does the clustering of similar values appear to represent a "hotspot" or a "coldspot"; or Negative SA, and if so does the spatial outlier appear to be surrounded by higher or lower values than its own; and last, whether there is any statistical significance associated with the type of SA present.

Performing these additional Local SA analyses does add more work, but given the limited amount of information that can be obtained from doing just a single KDE, I would recommend performing both the G_i^* and I_i when working with weighted data containing only coral disease

information. Additionally, while it is possible that some form of ecological analysis could be performed on case only datasets, I would not recommend spending too much time on such analyses. This is because without information on the underlying coral population at risk, there is no way of knowing whether or not the clusters detected by the Tier 2 analyses represent real disease clusters or are simply reflections of density changes in the underlying coral population. Thus, the validity of any correlations found during a Tier 3 ecologic analysis using case-only data would be questionable at best.

7.2.2 Recommendations When Both Coral Disease and Population Data are Available

When a given coral health dataset contains data on both the diseased corals as well as the underlying coral population at risk, there are not only more options and flexibility in the types of analysis the researcher choses to perform, but the results of these analyses are considered far more robust and reliable when compared to the case-only versions of the same analysis. **Table 7.3** provides the types of spatial analysis I recommend performing on either non-weighted or weighted datasets containing both numerator and denominator data. In which the "non-weighted data" column refers to datasets that contain presence/absence type data providing only the locations in which a given coral disease was found to be either present (case locations) or absent (control locations), but with no indication of how many diseased and/or non-diseased coral colonies were present at each location. Conversely, the "weighted data" column refers to datasets that contain disease prevalence data providing not only the locations in which a given type of coral was found, but also included on the total number of diseased and non-diseased coral colonies present at each location.

As with the case data, I recommend that the researcher begin with Tier 1, by creating a map to visualize the spatial distribution of the diseased and non-diseased corals. Non-weighted

data are best represented using point-based maps, in which all the locations are represented using the same cartographic symbol, but use different colors to differentiate between the case and control locations (see **Figure 7.1A**). Conversely, when weighted data are available, I recommend using polygon-based maps, in which each polygon represents the location where the given type of coral was found. The weighted data can be visualized as either one Case Prevalence map (**Figure 7.1B**), or as separate Case and Control density maps (**Figure 7.1C**), depending on the type of information the researcher is interested in.

Next, the researcher should move on to Tier 2's recommended methods for detecting and analyzing disease clusters (listed in **Table 7.3**). First *Ripley's K* analyses should be performed on the non-weighted and (if possible) weighted versions of the coral disease and underlying coral population data. The same normalization procedure that was described for the case-only data should be repeated for the weighted and/or non-weighted versions of the disease and population data. The same *Ripley's K* plots described for the case-only data also apply to the case data here, and can also be performed on the population data. The major advantage of having both the coral disease and population data are that the Difference function (D) can be used to examine the spatial distribution of the coral disease data after the spatial distribution of the underlying population has been accounted for. For the non-weighted coral data this is done by subtracting the *Normalized Observed K* values for the coral population locations from the *Normalized Observed K* values for the coral disease locations, such that $D=[L(d)_{\text{Cases}} - d] - [L(d)_{\text{Population}} - d]$. This D is then plotted against the tested distances (d), and using the *Normalized CI* for the Population data, resulting in a graph similar to the one shown in **Figure 6.7A** on page 137. For weighted coral data, this same process would then be repeated using the *Normalized Observed K* values for the weighted disease and population data, resulting in a graph similar to **Figure 6.7B** on page 137.

Table 7.3 Recommended Types of Spatial Analyses when Both Coral Disease and Underlying Coral Population Data are Available.

Geospatial Coral Epizootiology Tiers & ESDA Types	**non-weighted data**	**weighted** (locations weighted by # of individuals)
(1) Disease Mapping & Visualization[1]	Case & Control locations are in the same map	Case & Control densities are in separate maps
• Point-based maps (using circular, dot-shaped symbols)	Dots of uniform size in 2 different colors (see **Fig. 7.1A**)	—
• Polygon-based maps (using Thiessen Polygons)	—	Color-ramped shading based on weight (see **Fig. 7.1C**) or Show weighted Case & Population data in the same map by creating a Case Prevalence Map (see **Fig. 7.1B**)
(2) Detection & Analysis of Disease Clusters		
(2_A) General Disease Clustering (no spatial output)		
2. Point Pattern Analysis (PPA) Distance Statistics[1]	(see **Figures 6.7A** & **6.6A**,**C**, respectfully)	(see **Figures 6.7B**, **6.6B**,**D**, **6.8A**,**B**, & **6.9A**,**B**, respectfully)
• ***Ripley's K*** $[L(d)]$ [1]	Create plots of distance (X axis) vs. the following:	Create plots of distance (X axis) vs. the following:
Separate estimates are done for the Case & Population data	$\boldsymbol{D = [L(d)_{Cases} - d] - [L(d)_{Population} - d]}$ **&** $\boldsymbol{CI_{Population} - d}$	$\boldsymbol{D = [L(d)_{wCases} - d] - [L(d)_{wPopulation} - d]}$ **&** $\boldsymbol{CI_{wPopulation} - d}$
Results *Ripley's K* analyses are plotted in excel using	$L(d)_{Cases} - d$ & $CI_{Cases} - d$	$L(d)_{wCases} - d$ & $CI_{wCases} - d$
normalized *Observed K* values $[L(d) - d]$,	$L(d)_{Population} - d$ & $CI_{Population} - d$	$L(d)_{wPopulation} - d$ & $CI_{wPopulation} - d$
normalized upper & lower Confidence Intervals $[CI - d]$,		$L(d)_{wCases} - d$ & $CI_{Cases} - d$
normalized *Expected K* values of $[d - d]$ or $y = 0$		$L(d)_{wPopulation} - d$ & $CI_{Population} - d$
		$L(d)_{Cases} - d$ & $CI_{wCases} - d$
		$L(d)_{Population} - d$ & $CI_{wPopulation} - d$
(2_B) Specific Disease Clustering (has spatial output)		
3. Spatial Filtering & Smoothing[2]	(see **Figure 7.2B$_2$**)	(see **Figure 7.2B$_3$**)
• Dual Fixed Distance KDEs with Monte Carlo simulationsusing bandwidths calculated using the h_{opt} statistic	Performed simultaneously on the locations of the Case (numerator) & Population (denominator) data	Performed simultaneously on the weighted locations of the Case (numerator) & Population (denominator) data
4. Spatial Scan Statistics[3]	Bernoulli probability model compares the spatial distribution of the Case locations (disease present) to the distribution of the Control locations (disease absent) (see the Bernoulli model's NGO results in **Figure 6.17**)	Poisson probability model compares the spatial distribution of the Case locations weighted by # of individual cases present at each location to the spatial distribution of the Population locations weighted by the total # of individuals (see the Poisson model's NGO results in **Figure 6.17**)
5. Spatial Autocorrelation (SA) [1]		Thiessien Polygons generated from Population Locations
○ **Local SA analyses**	—	Performed separately on the Case, Control, & prevalence (case individuals/total individuals). The resulting index values, cluster types, & cluster significance (p-values) for each analysis are then mapped (see **Figure 6.18 A$_{2\text{-}3}$**, **B$_{2\text{-}3}$**, & **C$_{2\text{-}3}$**)
• *Getis-Ord Local G* (G_i^*)		
• *Local Moran's I* (I_i)		
(3) Disease Modeling, Prediction, & Ecological Analysis		
6. Spatial Regression[1]		
• *Ordinary Least Squares* (*OLS*) Regression in conjunction with *Geographically Weighted Regression* (*GWR*)	—	Used to compare the spatial distribution of the weighted Case data to the spatial distribution of other variables

Superscript numbers indicate which software to perform the analysis in: 1 = *ArcMap* 2 = *DMAP* 3 = *SaTScan*

While the *Ripley's K* plots based only on case or population data are useful and can be used to test various hypotheses, it is the plots based on the Difference Function that offer the most important insight into the spatial nature of a given coral disease outbreak. When time is an issue I recommend skipping the various versions of the *Normalized Ripley's K* data in which the case and population data are plotted separately. Instead, just use the difference function to create one plot that shows the spatial distribution of the coral disease data after the spatial distribution of the underlying population has been accounted for.

Once the researcher is satisfied with the results of their General Disease Clustering analyses, they should move on to Specific Disease Clustering methods in order to see where exactly these disease clusters are occurring. I recommend starting with performing a dual fixed-distance kernel density estimate (KDE) in *DMAP,* using an Optimized Bandwidth (calculated using the h_{opt} statistic and the locations from the population data) and 1,000 Monte Carlo simulations. The only difference between performing the dual KDE on the non-weighted and weighted datasets is that the input files used for the numerator (diseased locations) and denominator (population locations) for the non-weighted data will all have a weight of 1 in the weight column. Whereas, the weighted data will have the number of diseased colonies and total number of colonies for each location in the weight columns of the numerator and denominator input files, respectively. By displaying the *DMAP* results in *ArcMap* using the technique described in the legend of **Figure 7.2**, the researcher will not only have a map showing continuous surface of coral disease prevalence (also referred to as disease clustering rates), but they will also be able to see which areas have statistically significant prevalence rates.

I also recommend performing both the G_i^* and I_i local SA analyses on the weighted version of the case prevalence data (the weighted case data divided by the weighted population data, see **Figure 6.18C$_{2\text{-}3}$**). If time permits, I would also recommend performing these same

local SA analyses on the weighted case data and the weighted control data (see **Figure 6.18A$_{2\text{-}3}$** and **B$_{2\text{-}3}$**, respectively on page 159) so that additional information can be obtained about the spatial nature of both the diseased corals and the non-diseased corals. Or, alternatively, perform the local SA analyses on the weighted case and weighted population data, so that the spatial nature of the diseased corals can be compared to that of the underlying coral population at risk of the given disease.

However, if a faster and/or simpler method of identifying the locations of statistically significant disease clusters is needed, than I recommend performing a purely spatial analysis in *SaTScan*, using either the Bernouli probability model on non-weighted case (locations with disease present) and control (locations with disease absent) data, or the Poisson probability model on weighted case and population data. *SaTScan*'s default cluster detection setting of "No Geographical Overlap" (NGO) will most likely suffice for most coral studies. However, as was mentioned earlier in this chapter, I do not recommend relying exclusively on *SaTScan* for detecting & analyzing the locations of coral disease clusters unless it is absolutely necessary, as both *DMAP*'s dual KDEs and *ArcMap*'s local spatial autocorrelation (SA) analyses provide a far more detailed and accurate representation of the spatial nature of diseases.

Last, if the given coral dataset contains weighted data with at least several hundred diseased locations, as well as information on other variables which could potentially impact the distribution of the given disease, then I also recommend spending some time trying to develop models that will not only explain the spatial distribution of a given disease outbreak, but also potentially be used to predict future outbreaks. This can be done using either spatial regression methods or other types of ecologic analyses. The results from such Tier 3 methods have the potential to provide the most meaningful insight into current and future coral disease outbreaks. However, these methods are also highly complex, requiring not only a large amount of data, but

also a great deal of time to develop a well specified model through essentially countless rounds of trial and error trying to see which combinations of variables best explain the given disease distribution.

Table 7.4 Provides a summary of all of the different types of spatial analysis I recommend based on the types of coral health data available. In order to show how the correct use and application of geospatial techniques have the ability to greatly enhance our understanding of coral health. The following chapter applies some of the above recommended analytical methods to data from an actual coral disease outbreak.

Table 7.4 Summary of the Recommended Types of Spatial Analyses based on the Types of Coral Health Data Available.

Geospatial Coral Epizootiology Tiers & ESDA Types	Case (diseased)	Control (non-diseased)	Population (case + control)	Case & Control	Case & Population	Temporal Data	Other Attributes
(1) Disease Mapping & Visualization							
• Point-based maps (using circular, dot-shaped symbols)[1]	nw	nw	nw	nw	nw	—	—
• Polygon-based maps (using Thiessen Polygons based on point locations)[1]	w	w	w	nw & w	w (prevalence map)	—	—
(2) Detection & Analysis of Disease Clusters							
(2_A) General Disease Clustering (global statistic with no spatial output)							
○ Point Pattern Analysis (PPA) Distance Statistics							
• *Ripley's K* [1]	nw & w	nw & w	nw & w	—	—	—	—
(2_B) Specific Disease Clustering (local statistics with spatial output)							
○ Spatial Kernel Density Estimates (KDE)							
• Single Adaptive Distance KDEs [2]	nw & w	nw & w	nw & w	—	—	—	—
• Dual Fixed Distance KDEs with Monte Carlo simulations[3]	—	—	—	—	nw & w	—	—
○ Spatial Scan Statistic[4]							
• Bernoulli Model & default criteria for reporting secondary clusters	—	—	—	nw	—	—	—
• Poisson Model & default criteria for reporting secondary clusters	—	—	—	—	w	—	—
○ Local Spatial Autocorrelation (SA) Analyses[1]							
• *Getis-Ord Local G* (G_i^*)	w	w	w	—	w (prevalence map)	—	—
• *Local Moran's I* (I_i)	w	w	w	—	w (prevalence map)	—	—
(3) Disease Modeling, Prediction, & Ecological Analysis							
○ Space-Time Scan Statistic & associated probability models[4]	—	—	—	nw or w	nw or w	nw or w	—
○ Temporal Scan Statistic & associated probability models[4]	—	—	—	nw or w	nw or w	nw or w	—
○ Spatial Regression[1]							
• *Ordinary Least Squares* (*OLS*) Regression, in conjunction with	w	w	w	w	w	w	w
• *Geographically Weighted Regression* (*GWR*)	w	w	w	w	w	w	w

nw = non-weighted data (presence/absence by location)
w = weighted data (locations weighted by the # of individuals present at that location)

Superscript numbers indicate which software to perform the analysis in:
1 = *ArcMap* 2 = *CrimeStat* 3 = *DMAP* 4 = *SaTScan*

Chapter 8. Evaluating Patterns of a White-Band Disease (WBD) Outbreak in *Acropora palmata* Using Spatial Analysis: A Comparison of Transect and Colony Clustering[1]

"Spatial pattern analysis can document the scale of the disease processes and allow the testing of hypotheses about mechanisms of transmission" – Jolles et al. (2002)

8.1 Introduction

Over the past three decades, the incidence of coral disease has increased from sparse, localized sightings, to an apparent panzootic, as disease sightings have become commonplace among the world's reef systems. Since the first documented cases of coral disease in the late 1960s and early 1970s (Squires 1965; Antonius 1973,1981; Gladfelter 1982), scientists have been working to identify causes of these diseases (Harvell et al. 2007; Rosenberg et al. 2007a); however, progress has been slowed by the complexity of coral ecosystems and anthropogenic influences on these systems (Harvell et al. 2002; Kinlan and Gaines 2003; McCallum et al. 2003; Harvell et al. 2004; Sutherland et al. 2004; Ainsworth et al. 2007b; Harvell et al. 2007; Rosenberg et al. 2007a; Bourne et al. 2009; Johnson et al. 2010; Williams et al. 2010). Given the corresponding increase in human population pressure during this time period, it has been suggested that anthropogenic related stressors are contributing to, if not directly causing, coral disease outbreaks (Daszak et al. 2000,2001; Western 2001; Harvell 2004; Hoegh-Guldberg 2004; Sutherland et al. 2004; Aronson et al. 2005; Harvell et al. 2007; Lesser 2007; Baskett et al. 2010). While correlations between anthropogenic stressors and disease frequencies have been seen for quite some time (Green and Bruckner 2000; Daszak et al. 2001; Bruno et al. 2003; Gardner et al. 2003; Buddemeier et al. 2004; Johnson et al. 2010), it was only recently that direct

[1] This chapter has been reprinted permission from *PLoS ONE* (see **Appendix D**) with slight modifications. For the original publication please see Lentz JA, Blackburn JK, Curtis AJ (2011) Evaluating Patterns of a White-Band Disease (WBD) Outbreak in *Acropora palmata* Using Spatial Analysis: A Comparison of Transect and Colony Clustering. PLoS one 6:e21830

experimental evidence was able to actually show how anthropogenic stress factors (such as climate change, water pollution, and overfishing) were directly contributing to coral disease (Bruno et al. 2003,2007; Rosenberg et al. 2007a; Ainsworth and Hoegh-Guldberg 2009).

While coral diseases are occurring globally, their incidence appears to be the most severe in the Caribbean (Porter and Meier 1992; Grigg 1994; Hubbard et al. 1994; Hughes 1994; Jackson 1997,2001; Cooney et al. 2002; Bruno et al. 2003; Sutherland et al. 2004; Aronson and Precht 2006; Ainsworth et al. 2007b; Carpenter et al. 2008; Bourne et al. 2009; Miller et al. 2009). Over the past few decades reports show that disease is responsible for a roughly 80% loss in Caribbean coral cover (Gardner et al. 2003; Nowak 2004; Wapnick et al. 2004). Within the Caribbean, the *Acropora* coral genus appears to have been the hardest hit by disease, with *A. palmata* showing a 90-95% decline (Aronson and Precht 2001; Precht et al. 2002; Vollmer and Kline 2008; Bourne et al. 2009) and *A. cervicornis* populations collapsing across the region (Harvell et al. 2001; Bythell et al. 2002; Precht et al. 2002; Wapnick et al. 2004), causing them to be the first corals in history to be listed as "threatened" under the United States Endangered Species Act .

In 1977, shortly after the first documented coral disease, Black-Band Disease (BBD) (Antonius 1973,1981), a second "band" disease was also discovered in the Caribbean (Gladfelter 1982; Aronson and Precht 2001). This new White-Band Disease (WBD) has since been found to occur nearly worldwide in coral-supporting latitudes, ranging from the western Atlantic to the Red Sea, South Pacific, and Arabian Sea (Green and Bruckner 2000; Bythell et al. 2002). However, to date WBD has only been found to occur in the genus *Acropora* (Green and Bruckner 2000). Despite the well-known phenomenon of WBD, far less is known about its etiology, such as specific pathogen or pathogenic communities (e.g. BBD microbial communities), transmission dynamics or routes of infection (Casas et al. 2004; Sutherland et al.

2004; Aronson et al. 2005; Williams and Miller 2005; Pantos and Bythell 2006; Zvuloni et al. 2009).

WBD is visually identified by a white band of tissue separating the living tissue from the dead tissue (Gladfelter 1982). The specifics of this disease's appearance are important to note because all too often signs of predation and bleaching are mistaken for WBD (Ginsburg 2000). As the disease band moves, coral tissue is found peeling or sloughing off where the white band is, leaving behind exposed white skeleton (Gladfelter 1982; Richardson 1998; Williams and Miller 2005). In most cases, the coral skeleton does not remain bare for long, as the void is replaced by rapidly colonizing filamentous algae (Richardson 1998). This, combined with its rapid rate of spread, as much as $2.06 cm^2$/day, enables WBD to be the only known coral disease able to drastically change the structure and composition of reefs (Precht et al. 2002).

While BBD has been confirmed to be associated with a community of bacteria (Cooney et al. 2002), this has not been confirmed for WBD (Aronson and Precht 2001; Richardson et al. 2001; Bythell et al. 2002) or Yellow-Band Syndrome (YBS; Foley et al. 2005). However, it is often presumed that WBD is caused by a bacterial infection (Antonius 1981; Gladfelter 1982; Peters et al. 1983; Aronson and Precht 2001; Mayor et al. 2006). To date no pathogen has been isolated in pure culture, nor causation proven (Casas et al. 2004; Pantos and Bythell 2006; Vollmer and Kline 2008). However, the repeated findings of distinct differences between the bacterial communities present in healthy versus diseased tissue has lead recent studies to suggest that bacteria are more than just opportunistic invaders but rather appear to be associated with the disease – if not directly responsible for it (Casas et al. 2004; Pantos and Bythell 2006; Vollmer and Kline 2008). Some studies have proposed that WBD may not be pathogen-induced, but rather a biochemical response to some type of coral trauma, in essence a "shut-down-reaction" (Antonius 1981; Richardson 1998). Studies show that the frequency and severity of WBD

outbreaks over the past 30 years are unprecedented on a paleontological scale, leading many to speculate that anthropogenic stressors are directly associated with the disease, although to date no direct evidence of this reported (Richardson 1998; Aronson and Precht 2001; Precht et al. 2002; Gardner et al. 2003). The stressors that have been implicated include both regional stressors which are caused by the increasing human population levels coupled with anthropogenically driven climate change, as well as local stressors (such as over fishing, sedimentation, habitat destruction, etc.). However, proving that WBD is linked to any of these stressors is quite difficult without a known pathogen or etiologic agent, if one even exists. Further, if WBD is not pathogen induced, but rather the manifestation of the declining health of corals due to increased stress, then theoretically a diseased state could be brought upon by increases in one stressor or small to moderate increases in multiple stressors; in which case the stressors involved would likely vary from case to case.

While there is debate over the causes of WBD, as well as the extent and severity of disease-related mortality in *Acropora*, studies increasingly are showing that virtually all areas of the Caribbean are at risk of degradation (Precht et al. 2002; Gardner et al. 2003). By 1982, Tague Bay (located just north of St. Croix, see **Figure 8.1**), where Gladfelter first identified WBD in 1977, had lost about 50% of its *Acropora* population (both the shallow occurring *A. palmata* and the deeper occurring *A. cevicornis*). Within five years as much as 95% of the original *Acropora* population had died (Precht et al. 2002; Williams and Miller 2005). The decline in *Acropora* populations is of particular importance because the genus is known for developing the reef framework (Shinn 1963; Zubillaga et al. 2008), as well as for providing habitat critical to the support of diverse reef fish populations (Lirman 1999) and other organisms that contribute to the productivity and overall health of the reef (Aronson and Precht 1997,2001; Jackson et al. 2001; Precht et al. 2002).

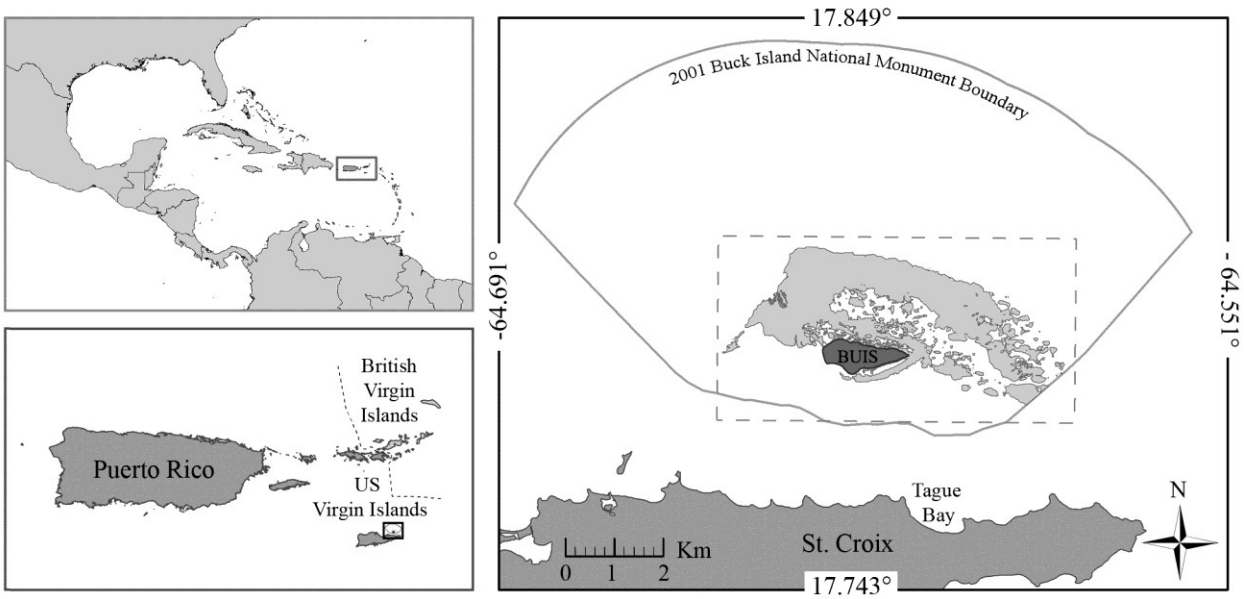


Figure 8.1 The study area. Buck Island (BUIS) Reef National Monument, located just north of the island of St. Croix, US Virgin Islands (USVI). Mayor et al.'s (2006) study area is delineated by the light grey area surrounding BUIS, consisting primarily of hard-bottom substrate less than 10*m* deep. The extent of the grid surface used in the DMAP analysis is depicted by the dashed rectangle surrounding the study area.

Over the last decade there has been increased recognition that geography plays an important role in coral diseases, marked in large part by the growing number of studies that employ geographic information systems (GIS) technologies and spatial statistics (Jolles et al. 2002; Foley et al. 2005; Zvuloni et al. 2009); though to date, relatively few studies have directly analyzed the spatial patterns of diseases in reef communities. Jolles et al. (2002) provides a key approach to the application of spatial statistics to explore spatial patterns of aspergillosis (a diseased caused by the fungus *Aspergillus syndowii*) in sea fans to test hypotheses of transmission and infection.

Jolles et al. (2002) employed the *Ripley's K* statistic, a global measure of spatial autocorrelation, to describe the spatial patterns of disease in sea fans of various sizes and from multiple sites with regard to the underlying sea fan population. Specifically, they were interested in determining whether the spatial distribution of diseased sea fans was clustered, dispersed, or

random at each of the different distance scales tested, and how this diseased distribution compared to the spatial distribution of the underlying sea fan population. By doing this they were able to not only quantify the geographic scale of the disease outbreak, but they were also able to test hypotheses regarding the secondary transmission of *A. syndowii*. Their results showed that where disease prevalence was low, the disease appeared to have a random spatial distribution; which might indicate that the disease was being transmitted by terrestrial sources (such as soil runoff or airborne dust). Conversely, they found that where disease prevalence was high there would be a statistically significant spatial aggregation (cluster) of aspergillosis; which would be more indicative of secondary transmission of the disease through either direct contact (sea fan to sea fan, or through a vector such as fish or snails) or through the water column.

More recently, a similar approach was used to study the spatio-temporal patterns of BBD in order to assess possible disease transmission mechanisms (Zvuloni et al. 2009). Specifically, they used the *Ripley's K* statistic in both their spatial and spatio-temporal analyses to infer transmission patterns and to calculate epidemiologic parameters, such as the basic reproductive number (R_0). Their study found that BBD was spatially clustered (though not to a statistically significant extent) and that as the peak disease season was approached the size of these clusters would increase. The temporal nature of their study enabled them to track disease spread throughout their study area. Over the course of their two year study, they found that newly infected corals were often in close proximity to (or even in direct contact with) already infected corals, indicating that BBD was likely being spread through the water column and by direct contact with infected individuals. Ultimately, they reached a similar conclusion as Jolles et al. (2002), stating that the presence of disease clusters were the "hallmark signature for the presence of localized transmission dynamics" (page 9, Zvuloni et al. 2009).

The GIS and spatial analytical methods employed by Jolles et al. (2002) and Zvuloni et al. (2009) facilitated a better understanding of the etiologies of their respective diseases by examining the spatial disease distribution, and testing hypotheses regarding the mode of transmission and infection. However, it is important to note that both of these studies were based on diseases in which the infectious agent had already been identified. Unfortunately, this is not the case for most coral diseases.

A novel study by Foley et al. (2005) used GIS and spatial analysis (specifically the *Ripley K* function) to study the spatial distribution of YBS in an effort to infer causation from spatial patterns of disease. Their results revealed that while the underlying population of susceptible corals (*Montastrea annularis*) appeared to be strongly spatially clustered, the distribution of *M. annularis* with YBS was less clustered and more dispersed (Foley et al. 2005). Those results were consistent with hypothesized etiologies in which near shore pathogens or toxins were either directly introducing YBS or indirectly leading to YBS by increasing host susceptibility (Foley et al. 2005). They postulated that the lack of disease clustering in a population in which the individuals show a strong spatial aggregation, may indicate that the close proximity of the corals may decrease the risk of infection by creating physical barriers which would inhibit the transmission of the disease agent or toxins (Foley et al. 2005).

Following the rationale of Foley et al. (2005), this paper employs spatial statistics in an effort to characterize the patterns of WBD in *A. palmata* colonies from a 2004 outbreak in the reef system around Buck Island National Monument, St. Croix, US Virgin Islands (USVI, see **Figure 8.1**) using data from Mayor et al. (2006). In an effort to characterize the prevalence of WBD and the extent of elkhorn coral damage from disease and hurricane damage, Mayor et al. (2006) initiated an intensive sampling effort to map and count colonies of *A. palmata.* That initial study documented a prevalence of ~3 % WBD across colonies and suggested that it may

still pose a threat to the Buck Island reef community. This study employs the *Ripley's K* statistic, and a spatial filtering method to identify local spatial clusters of disease and discusses those in the context of possible causative agents or reef trauma that may assist in the ultimate determination of WBD causation.

8.2 Materials and Methods

Spatial analyses were performed on data provided by the US National Park Service. The dataset was originally compiled in a study examining the distribution and abundance of *A. palmata*, and the prevalence of WBD around Buck Island (BUIS) following a 2004 outbreak (Mayor et al. 2006). In order to facilitate data collection, the original survey evaluated habitats favorable for *A. palmata*, limiting the survey region to hard-bottom areas less than 10*m* deep (depicted as the shaded region around BUIS in **Figure 8.1**). A total of 617 locations were randomly selected for 25*m* by 10*m* transect surveys. Of those transects, 375 contained *A. palmata* colonies. Following the original case definition of Mayor et al. (2006), "Elkhorn colonies were considered infected with WBD if they had narrow white bands of exposed skeleton, circling completely around the coral branches, bordered on the upper side by live tissue and on the lower side by dead skeleton covered with algae" (page 240). Of those 375 original transects 44 contained evidence of WBD.

Spatial locations were recorded for each transect and not for each individual coral colony, although each transect location had a total number of colonies associated with it. To test for potential differences in WBD prevalence estimates and spatial patterns between those two scales, two subsets of the spatial data were developed. The first subset consisted of "transect-level" data of WBD presence or absence. While, the second subset consisted of "colony-level" data in which each transect location was weighted by the number of *A. palmata* colonies (both with and without WBD) present at that location (see **Figure 8.2**).

8.2.1 Spatial Autocorrelation Methods

The *Ripley's K* statistic was employed in *ArcGIS 9.3.1* to examine the extent of spatial dependence (the clustering or dispersion of corals) across several distances. This statistic was calculated using the following linear transformation of the K-function:

$$L(d) = \sqrt{\frac{A\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} k(i,j)}{\pi n(n-1)}}$$

where n is the total number of transect locations, k is used to weight the transect location (i,j) by the number of *A. palmata* colonies within the given transect, A is the study area, and d is the distance over which the spatial autocorrelation is being tested. The distance, d, was calculated from 0 to 2,500m in 50m bins for corals with WBD present, corals without WBD present, and for the underlying coral population for both the transect-level and colony-level subsets. Note no weight (k) was included in the transect-level analyses. A total of six analyses were conducted. For each, 99 permutations were run resulting in a 99% (or 0.01) confidence interval (CI) envelope for the *Observed K*s. The resultant *Observed* and *Expected K* values (*L(d)* and d, respectively) were plotted against the tested distances for each of the 6 analyses. The *Expected K* values represent the null distribution of complete spatial randomness (CSR), also known as the "Poisson distribution." The plotted *Expected K* values act as the benchmark used to test the spatial distribution of the *Observed K*s against the null distribution of CSR. The *Observed K*s that fall along this line are considered to have a spatially random distribution, while anything that lies above this line is considered to have a more aggregated spatial distribution and anything that falls below this line is considered to have a more dispersed spatial distribution. The CI envelope is used to determine whether or not the observed spatial pattern is statistically significant (p = 0.01), with no significance associated with the spatial distributions of *Observed K*s within this

envelope. The observed distribution is considered to have significant clustering when the values lie above the upper CI; conversely, values that lie below the lower CI are considered to be significantly dispersed.

The difference function (*D*) was used to examine the spatial distribution of WBD with respect to underlying environmental heterogeneity caused by the presence of the underlying coral population. To do this the *Normalized K* values from the underlying population were subtracted from those of the WBD corals so that I would be able to assess to what extent the spatial distributions of WBD depicted by the homogeneous analyses (**Appendix C**) were caused by the disease itself, rather than the natural background variation in the *A. palmata* population (**Figure 8.3**). The resulting Disease-Population difference function was quite similar to the design of the *Ripley's K* function used by Jolles et al. (2002) in which they set their null distribution equal to that of the underlying population of susceptible corals and then plotted K-K_{null} against distance.

8.2.2 Spatial Filtering Methods

The *Disease Mapping and Analysis Program* (*DMAP*, available for download at *http://www.uiowa.edu/~gishlth/DMAP/*) was used to employ a spatial filter to smooth prevalence estimates and then identify statistically significant increased prevalence using Monte Carlo simulations (Rushton and Lolonis 1996; Rushton et al. 1996; Curtis and Leitner 2006). These prevalence estimates are spatially explicit and represent clusters on the mapped surface. *DMAP* was used to construct WBD prevalence surfaces for both data subsets.

DMAP analyses require a rectangular gridded surface that encompassed the entire study area. The grid was defined in the northwest by lat/long coordinates of 17.809ºN, -64.648ºW, and in the south-east by 17.775ºS, -64.579ºE, respectively, with a $50m^2$ grid cell size (see dashed rectangle in **Figure 8.1**). Grid cell size was chosen based on the scale of the analysis and size of the study area. The size of the grid cell is important because it defines the scale of identified

cluster patterns, if the grid cells are too small the interpolation will become jagged, while an excessively large grid cell will lack resolution in delineating clusters.

All point level data are aggregated to a filter centered on each grid intersection point. In *DMAP* this filter is a circle with a user-defined radius. This filter is then applied to the numerator (transects containing *A. palmata* with WBD) and denominator (all transects containing *A. palmata*) data to calculate prevalence at each grid intersection. It is important to note that these filters must be large enough to cover multiple-grid intersections, allowing for points to be included in multiple prevalence calculations, and thus smoothing the estimated surface which eliminates hard (and often artificially defined) aggregation breaks. Once these local prevalence estimates have been calculated, a Monte Carlo simulation is employed to identify any areas with repeated prevalence estimates higher than expected from the simulations. The Monte Carlo simulation is based on the actual locations of transects containing *A. palmata* colonies; with a probability for each "healthy" individual becoming diseased. Probability was set as the prevalence of each of the transect and colony-level analyses, respectively. A Monte Carlo simulation re-creates this disease surface "*n*" times, creating a simulated distribution against which the actual disease surface is compared. If, for example, the prevalence in one filter is actually higher in 990 out of the 1,000 simulation runs, one can be 99% confident (equivalent to a *p*-value of 0.01) that the revealed prevalence, or hotspot, did not occur by chance alone. These hotspots are considered spatial clusters of WBD within the BUIS reef system.

As the method of WBD transmission is not currently known, nor the distance to which the pathogen or vector (if any) can viably travel, the spatial parameters used during the spatial analysis could not be based on the epidemiology of WBD. For this reason the optimized bandwidth (h_{opt}) statistic was used to estimate the size of the filter radius size based on the spatial

structure of the dataset. Following Fotheringham et al. (2000) the optimized bandwidth was calculated as:

$$h_{opt} = \left[\frac{2}{3n}\right]^{\frac{1}{4}} \sigma$$

where n = the sample size of transect locations (375) and σ = the standard distance or a measure of dispersion around the spatial mean of the transect locations. Standard distance was calculated in *ArcGIS 9.3.1* using the spatial statistics toolbox and a standard deviation of 1 (1688.2m). The resulting optimized bandwidth estimation (h_{opt} = 342.55m) was employed for *DMAP* analyses on both transect and colony-level data. Resultant hotspots were mapped in *ArcGIS 9.3.1* by rasterizing the *DMAP* output of the WBD prevalence estimates and overlaying probability value contours outlining disease clusters in which the of WBD prevalence estimates were statistically significant (p = 0.05).

8.3 Results

Given that WBD was found at 44 of the 375 transects surveyed, the estimated prevalence of WBD based on the transect-level data was $11.7\overline{3}\%$, suggesting that more than 10% of the transects reported diseased *A. palmata*. However, of the 2,492 colonies surveyed only 69 appeared to have WBD present, which results in a WBD prevalence of 2.77% based on the colony-level data. The mean number of *A. palmata* colonies with WBD absent per transect was 6.48 (min 1, max 40, 5.87 SD), which was very close to that of the overall mean, 6.65 (min 1, max 40, 5.99 SD); while, the mean number of *A. palmata* colonies with WBD present was much lower, 1.57 (min 1, max 6, 1.16 SD). The graph in **Figure 8.2A** illustrates the distribution of the number colonies with and without WBD present among the surveyed transects.

As transect- and colony-level analyses were performed on same coral dataset, it became clear how interpretations of the data would change based on the level of reporting (**Figure 8.2B**).

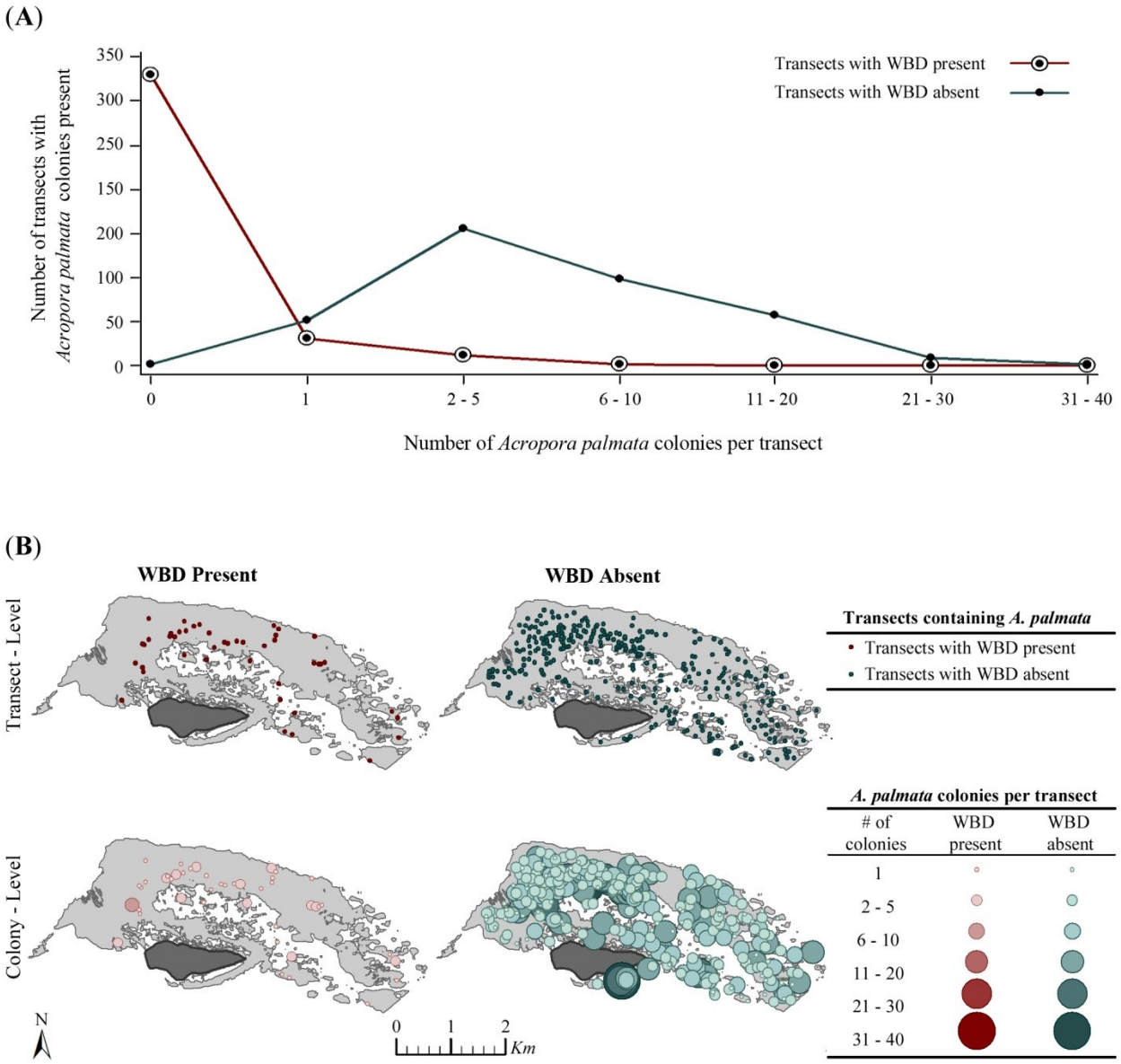


Figure 8.2 This figure visually depicts the differences between the transect- and colony-level versions of the dataset. (**A**) Colony densities (the number of colonies per transect) are plotted against the total number of transects with a given colony density, resulting in the cumulative frequency of the colony densities with and without white-band disease (WBD) present. (**B**) Circular symbols are used to indicate the locations of transects with and without WBD present, from the transect-level version of the dataset (top row). The colony-level dataset is depicted using a graduated symbol map in which the size and color of the symbols used to indicate the locations of each transect are scaled according to the number of colonies within that transect to depicts the colony-level dataset (bottom row).

The transect-level data represent the presence or absence of WBD for each transect, which was visually depicted in the top row of **Figure 8.2B** by circles indicating the locations of the 44 transects in which WBD was present (top left) and the 331 transects where no WBD was seen (top right). While, the second version of our dataset, consisted of the same geographic information (the transect locations); it included additional information about the disease-state of the individual colonies within each transect. The colony-level analysis of the dataset was visually depicted by circular-symbols in which the center of each circle indicated the transect location (**Figure 8.2B**), while the size and shade of the symbol were scaled to represent the number of colonies within each transect that either had WBD present (bottom left) or WBD absent (bottom right).

The most striking differences between the resultant spatial distributions of the transect- and colony-level versions of the dataset became apparent when the difference function (D) was used to examine the spatial patterning of WBD among the *A. palmata* coral populations (**Figure 8.3**). The presence/absence analysis of WBD at the transect-level (**Figure 8.3A**) revealed spatial aggregation in all transects containing WBD. No significant difference was detected between the aggregated distribution of transects with WBD present and the aggregated distribution of the 375 total transects, based on analysis done using distance thresholds between 1.25km and 1.50km; while the aggregation of WBD was found to be significantly more clustered at distance scales < 1.25km and significantly less clustered at distances > 1.50km than the clustered distribution of the underlying population. The weighted K function analysis of prevalence WBD at the colony-level (**Figure 8.3B**) revealed that colonies with WBD present had fairly random spatial distributions at distances < 2.1km, becoming more dispersed at distances >2.1km. However, when compared to the underlying population densities, the spatial distribution of the WBD colonies was significantly more dispersed than the clustered distribution of susceptible colonies.

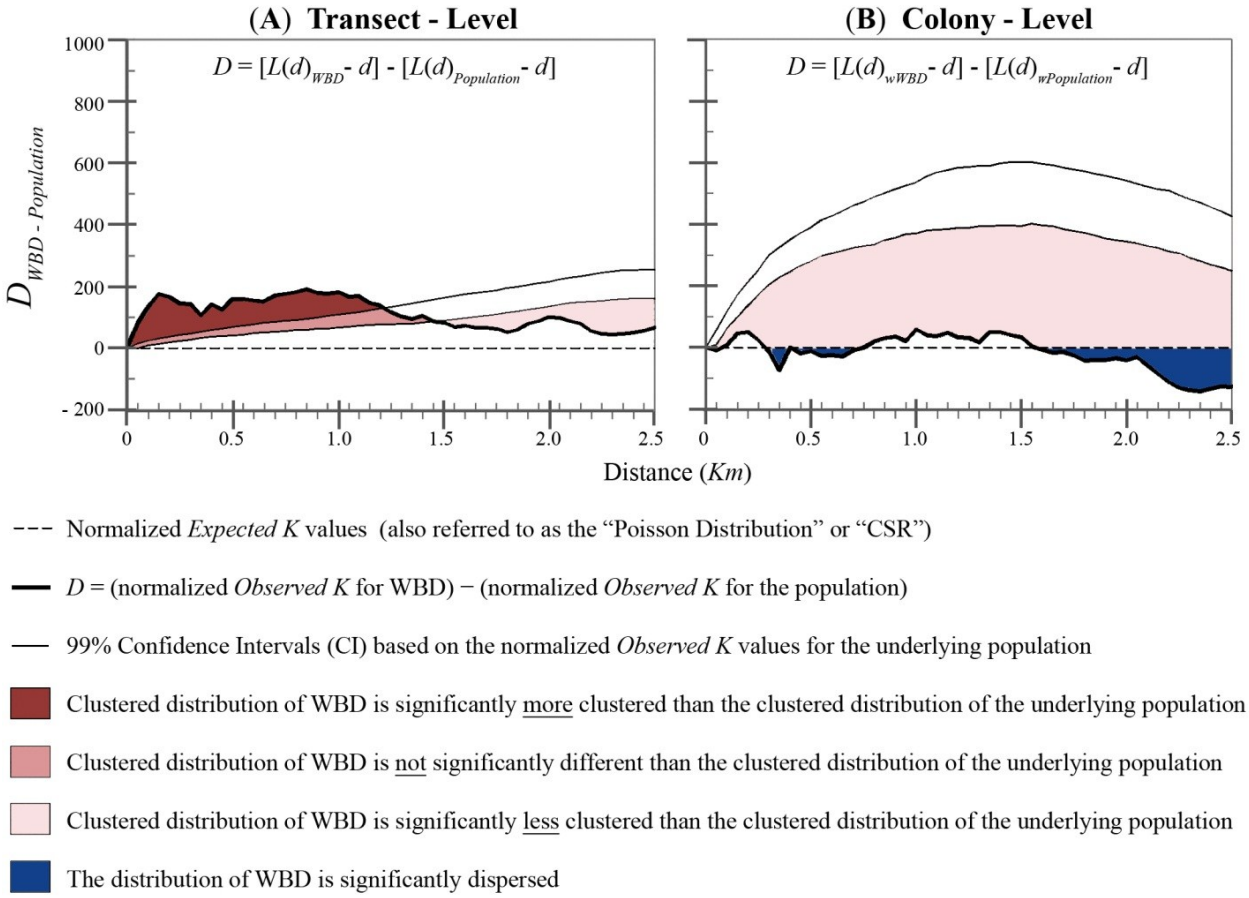


Figure 8.3 The results of the *Ripley's K* spatial autocorrelation analysis. Normalized *Ripley's K* plots were used to assess the spatial distribution of white-band disease (WBD) among *Acropora palmata* over a distance of 2.5*m*. Transect-level and colony-level versions of the *K* function were performed in order to compare the spatial distributions of WBD based on data analyzed at the (**A**) transect- and (**B**) colony-levels (respectively). In order to insure that the observed spatial distribution was reflecting the spatial nature of WBD, and not the spatial patterning of the underlying population, the transect and colony-level *Observed K* values for the underlying population were subtracted from the *Observed K*s of WBD at the transect- and colony-levels, respectively. The resulting *K* values for WBD were then plotted against distance. The spatial nature of WBD was then assessed by comparing these *K* values for WBD (thick line) to a spatially random (Poisson) distribution (dashed line at $y = 0$), in which WBD values above the Poisson distribution indicates WBD was clustered within the underlying population, while values below this line indicated WBD was more dispersed than the underlying population. The 99% confidence intervals (thin lines) generated from the *Observed K* values for the population were used to determine the statistical significance of distribution of WBD within the underlying population of susceptible corals.

Analyses using the DMAP spatial filter revealed significant spatial clustering at both spatial scales tested; however, it is interesting to note some differences in the distribution and size of clusters in each of the two experiments. A red line was used to show the exterior boundaries of areas in which the WBD prevalence estimates were predicted to be statistically significant ($p = 0.05$) based on 1000 Monte Carlo simulations (**Figure 8.4**).

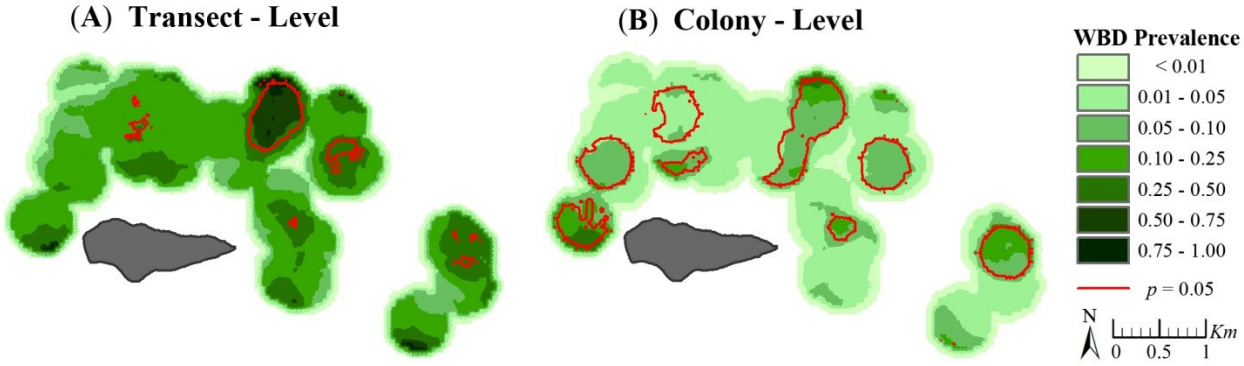


Figure 8.4 The results of the *Disease Mapping and Analysis Program* (*DMAP*) spatial filtering analysis. Comparing the difference between analyzing the coral dataset at the transect (**A**) verses colony-level (**B**) using *DMAP*. The following spatial parameters were used for both analyses: a $50m^2$ grid cell resolution; and a $342.55m$ filter radius, calculated using the Optimized Bandwidth (h_{opt}) estimation method. The prevalence of white-band disease (WBD) clustering are shown in green, with darker shades indicating increased prevalence. Areas with statistically significant clustering rates ($p \leq 0.05$), based on 1000 Monte Carlo simulations, are outlined in red. The numbers placed beside each significant clustering were used solely for identification purposes, and have no empirical value.

Overall, the transect-level analysis revealed relatively high WBD prevalence throughout the study area (indicated by the dark shades of green in **Figure 8.4A**), with approximately five areas with statistically significant WBD clustering. By comparing the spatial output to Mayor et al.'s (2006) dataset, we found that 36.4% of the transects with WBD present (containing 37.7% of the diseased colonies) were located within $100m$ of these five areas of significant disease clustering, with only 13.6% of the WBD transects (containing less than 12% of the total disease colonies) occurring inside one of the areas with significant WBD clustering.

The WBD clustering patterns revealed by the DMAP analysis of the colony-level dataset revealed dramatically different results. The prevalence of WBD was relatively low across the study area, with eight relatively large statistically significant areas of WBD clustering distributed fairly evenly throughout the study area (**Figure 8.4B**). When the areas of statistically significant WBD clustering were compared to our underlying dataset, we found that more than half of the transects and colonies with WBD (70.5% and 79.7%, respectively) were within 100*m* of one of the 8 significant clustering areas, of which 34.1% of the transects and 50.7% of the colonies were located inside one of the 8 areas.

The total area with significant WBD clustering based on the *DMAP* Monte Carlo analysis of the colony-level dataset was almost three times larger than the total clustering area based on *DMAP* analysis of the transect-level data (20.50km^2 and 7.35km^2, respectively), even though the WBD prevalence estimated at the transect level is more than four times higher than the prevalence estimated at the colony-level. The mean transect depth inside the significant clustering areas for the transect-level and colony-level datasets was 7.55*m* and 6.90*m*, respectively, compared to a mean transect depth of 5.87*m* for all transects surveyed within the study area.

8.4 Discussion

Despite being one of the first documented coral diseases, there is still little information available on the causative agent or specific environmental stressors that promote White-Band Disease (WBD; Aronson and Precht 2001; Casas et al. 2004; Sutherland et al. 2004; Aronson et al. 2005; Williams and Miller 2005; Pantos and Bythell 2006). As the search for causation continues, surveillance and proper documentation of the spatial patterns may inform etiology, and at the same time assist reef managers in allocating resources to tracking the disease.

These results show a clear difference between interpreting data at the transect verses colony-level (**Figures 8.2 – 8.4**). The disease surface produced by the transect-level analysis suggests that this was a severe, widespread WBD outbreak (indicated by the high WBD prevalence estimates throughout the study area, see the dark green areas of **Figure 8.4A**). Assuming that the disease is contagious and spreads from an initial location, one could hypothesize that the primary cluster areas identified by the transect-level analysis may be the origin of the outbreak, with cases spreading via the dominant direction of tidal flow, currents, prevailing winds, etc. This hypothesis could be tested with time-specific data on WBD occurrence or modeled with simulated data to determine if such a flow is feasible (Zvuloni et al. 2009). This would allow the development of a working spatial model for contagious spread based on reef morphology, water flow, and environmental conditions around the reef. However, testing this hypothesis was beyond the scope of this study, as Mayor et al.'s (2006) dataset we did not have a temporal component. In contrast, the disease surface produced by the colony-level analysis might indicate that a low-grade, broadly distributed WBD outbreak that might be the result of a ubiquitous stressor. In this way, the spatial resolution from each analysis can be used in a Modifiable Area Unit Problem (MAUP) framework to develop field studies and models designed to test these hypotheses to inform the etiology and subsequent pathogen surveys (Openshaw and Taylor 1979; Openshaw 1984).

The use of the spatial filtering approach here allowed me to evaluate the distribution of local clusters across the reef and identify specific hotspots of WBD for the 2004 data set. In this way, I can evaluate specific hydrological conditions, reef morphology, or environmental contamination (or microbial communities) that might influence specific regions of the reef that might now be acting globally across reef. While the use of *Ripley's K* by the seminal works Jolles et al. (2002) and Foley et al. (2005) provided insights in to the spatial pattern and scale of

the apsergillosis in sea fans and YBS in corals, respectively, the precise location of clusters must be inferred in those studies based on sampling strategy and reef location. The *Ripley's K* statistic is a global measure designed to determine the spatial scale at which clustering is present on the landscape, but it does not identify where on the landscape the clustering is occurring (Gatrell et al. 1996; Lancaster and Downes 2004; Marcon and Puech 2009; Bayard and Elphick 2010).

As did Jolles et al. (2002), Foley et al. (2005), and Zvuloni et al. (2009), this study directly accounted for the distribution of both infected and unaffected corals, allowing me to test and ultimately reject the hypothesis that clusters in WBD were simply reflections of the underlying coral density. The prevalence of WBD was much lower than the prevalence of aspergillosis in Jolles et al. (2002) study in which the mean prevalence among their 3 sites was 47.97%, whereas, the prevalence of WBD was only 2.77% and 11.73% based on the colony- and transect-level datasets respectively. Jolles et al. (2002) found significant clustering in areas of high disease prevalence. The *Ripley's K* results of our transect-level data (**Figure 8.3A**) support this, given that the WBD prevalence estimated at the transect-level was much higher than that of the colony-level, and there was the high degree of significant WBD clustering (compared to the aggregated distribution of the underlying transects) based on the transect-level data, whereas no significant WBD clustering was detected using the colony-level *Ripley's K* analysis of the colony-level data (**Figure 8.3B**). However, this does not appear to be the case when the results of the *DMAP* analyses were examined, as the colony-level data had a total significant clustering area almost three times larger than that of the transect-level data, but the WBD prevalence estimated at the transect level was more than four times greater than colony prevalence.

The low prevalence of WBD among *A. palmata* colonies, combined with the fairly random spatial distribution of WBD colonies shown in **Figure 8.3B**, might indicate that the disease is caused by either air and/or water-born direct transmission of the causative disease

agent from a terrestrial point of origin (Jolles et al. 2002). The rational being that corals "of equal size have equal chances of being hit by infectious material suspended in the water column" (page 2374, Jolles et al. 2002). The assumptions of this hypothesized mode of disease transmission were supported given that the overall distance between possible terrestrial-based contaminant sources and the locations of the *A. palmata* colonies was quite large compared to the significantly clustered spatial distribution among the susceptible colonies (Jolles et al. 2002). In addition, the dispersed WBD distributions might also indicate that the clustered coral population may offer protection from disease by providing physical barriers to the disease agents or toxins (Foley et al. 2005).

The presence of statistically significant areas of WBD clustering, as indicated by the *DMAP* analyses, does not necessarily conflict with the assumptions of this hypothesis, as the type of cluster analysis used to test this theory by the previously mentioned studies (i.e. the *Ripley's K* function) was based on a global statistic designed to quantify changes in spatial patterns at various distances. Instead, given the low WBD prevalence estimates and broad geographic distribution of the areas with statistically significant disease clusters identified by the *DMAP* analysis, the colony-level data could be used to support this hypothesis, suggesting that WBD might be the result of a ubiquitous stressor. In such a case, the areas of significant disease clustering, might indicate the presence of locally aggregated stress factors which might make the surrounding corals more vulnerable to infection (suggested by Jolles et al. 2002). This hypothesis could be tested by looking for correlations between areas with increased environmental risk factors and the areas of significant WBD clustering predicted by *DMAP* (or other types of spatial filtering analysis) in comparison to areas absent of disease in the study area. Conversely, WBD clusters may indicate the presence of diverse microbial organisms with different virulence levels, though the causative agent(s) and mechanism are not yet described.

Disease clustering could also be the result of genetic clustering of corals that are more susceptible to the disease. This possibility was ruled out by both Jolles et al. (2002) and Zvuloni et al. (2009) as genetic clustering was unlikely due to the reproductive nature of the corals in their studies (sea fans and massive corals respectively). *A. palmata* can reproduce both sexually via broadcast spawning (Szmant 1986; Baums et al. 2006) – which would make genetic clustering unlikely (Jolles et al. 2002; Zvuloni et al. 2009), and asexually through fragmentation (Highsmith 1982; Lirman 2000). Historically, *Acropora* relied on seasonal sexual reproduction to increase their population size and distribution, while using asexual fragmentation as a survival mechanism to rebound from storms or other physical damage (Highsmith 1982). Ultimately, one of the traits that had made *A. palmata* so resilient in the past may be a contributing factor to their decline, as the decrease in genetic diversity that tends to occur in populations dominated by fragmentation may cause the corals to be more susceptible to emergent epizootics (Bak 1983; Bruckner 2002a; Williams et al. 2008). In addition, when fragmentation occurs the corals have to devote their energy towards recovery instead of reproduction (Lirman 2000; Baums et al. 2006). The same appears to be true of stress in general for *Acropora*, as populations recovering from various high stress conditions show decreased (or the complete cessation of) sexual reproductive processes, as well as decreased survival of their asexual fragments (Williams et al. 2008). How long it takes for *A. palmata* to recover enough from fragmentation or other stresses to start spawning appears to vary by location (Baums et al. 2006; Williams et al. 2008). Lirman's (2000) study showed that "3 years after Hurricane Andrew, gametes were only present in large *A. palmata* colonies that had not experienced direct fragmentation during the storm. Neither those colonies that were damaged by the hurricane nor any of the hurricane-generated fragments had produced gametes at this time" (page 53). Additionally it appears that "colony fecundity is dependent on a coral's size and condition" (page 124, Grober-Dunsmore et al.

2006), which is a problem because stressors appear to disproportionately affect the larger colonies (Grober-Dunsmore et al. 2006).

Overall the combined low disease prevalence, limited number of (large) clusters, and wide distribution of statistically significant WBD clusters suggests WBD may "persist as a ubiquitous, chronic stress," as was suggested by Grober-Dunsmore et al. (2006) for the *A. palmata* in their study area (which surrounded the island of St. John, located just north of Buck Island in the USVI).

At present, many investigations examining spatial data concentrate (and more importantly, collect) only on the variable of interest. In the case of coral disease, this would be the location of the diseased coral. However, without similarly collected population data, it is impossible to know if the pattern revealed by the analysis is a disease "hotspot," or simply indicative of locations with higher densities of coral; i.e., ceteris paribus, the more coral there is, the more diseased corals are likely to be found. However, a counter problem of weighting a transect by the number of colonies is – exactly where do colony boundaries occur? It is possible to create an artificial hotspot by adding too many artificial boundaries. For these reasons, studies examining coral diseases should be done at as fine a spatial resolution as possible, with accurate and precise spatial measurements. This will have the added benefit of not only improving existing spatial investigations but opening the analysis to more sophisticated spatial inquiry.

Future studies should also examine each of these significant WBD clustering areas at both the geographic and microbial scales. In this way spatial regression models could be used to associate disease clusters with surrounding environmental factors, such as stressors (human population size, pollution, frequently visited tourist sites, etc.), and/or physical properties (surface currents, sea surface temperatures, wind direction, salinity, etc). Analyses at the microbial scale could test for similarities and differences in the histology and bacterial

communities between corals from each of the significant diseased clusters; as well as compare corals within significant disease clustering areas to those in non-significant diseased areas.

The analysis and mapping approach employed here can also be used to study the spatio-temporal changes in coral health by comparing changes in the position, size, and local prevalence rates of clusters and significant areas of coral bleaching and other coral diseases. Comparisons of the clustering of different types of diseases present in one location may also provide valuable insight into the continued decline in reef health worldwide. These spatial insights should provide valuable insights to both coral disease researchers and marine resource managers with information on the most vulnerable areas of the reefs.

Chapter 9. Synthesis and Conclusions

"Worldwide coral reef ecosystems have been transformed under the influence of direct and indirect effects of human activities (Bruno et al. 2007). Understanding the relationships between human activities and their ecological impacts and assessing the spatial distribution of these impacts are crucial steps in managing the use of coral reefs in a way that maximizes commercial and societal benefits while minimizing reef degradation."
– Selkoe et al. (2009)

9.1 Summary

Over the last few decades, recognition of the importance of geospatial information concerning health-related issues has increased substantially. In the coral community, this shift has been marked by a seemingly exponential increase in the number of studies reporting the importance of spatial pattern analysis in determining both the cause and transmission dynamics of various coral diseases (Real and McElhany 1996; Foley et al. 2005; Crowder et al. 2006; Grober-Dunsmore et al. 2006; Ritchie 2006; Selig et al. 2006; Jones et al. 2008; Selkoe et al. 2009; Sokolow 2009; Weil and Croquer 2009; Zvuloni et al. 2009; Eakin et al. 2010; Ruiz-Moreno et al. 2010; Selig et al. 2010; Maina et al. 2011; Pittman and Brown 2011). As part of this recognition, the recording of GPS coordinates has become standard protocol for many coral field studies (Ginsburg 2000; Jolles et al. 2002; Weil et al. 2002; Lang 2003; Willis et al. 2004; Foley et al. 2005; Grober-Dunsmore et al. 2006; Mayor et al. 2006; Zvuloni et al. 2009; ICRI/UNEP-WCMC 2010b; Pittman and Brown 2011).

9.1.1 Spatial Analysis of Coral Diseases

Despite this growing recognition of the importance of studying the spatial nature of both coral diseases and their environmental stressors, only a handful of studies have actually used spatial statistics to analyze their spatial data (Jolles et al. 2002; Foley et al. 2005; Zvuloni et al. 2009). Meanwhile, the remainder of the studies claiming to spatially analyze their coral disease data, either rely on visual examination of disease locations (Grober-Dunsmore et al. 2006; Mayor

et al. 2006; Selkoe et al. 2009), or use standard, linear statistics to analyze their spatial data (Selig et al. 2010; Maina et al. 2011). There are several problems with using traditional statistical techniques on spatial data. First, the very nature of these techniques treats the data as if all the points are occurring in the exact same location. Second, many of these statistics are based on underlying assumptions that the data has a normal, homogeneous distribution (Maina et al. 2011); which is an inappropriate assumption considering that both corals (as well as most living creatures) and environmental stressors almost always have heterogeneous spatial distributions (Harley et al. 2006; Ruiz-Moreno et al. 2010; Selig et al. 2010).

However, it should also be noted that the use of geospatial statistics should not be taken lightly, as different types of spatial analysis, as well as different parameter settings within each analysis, can produce noticeably different results. Consequently, poor selection or improper use of a given technique could lead to inaccurate representations of the spatial distribution, resulting in false interpretations of the disease. For this reason, a comprehensive review was done of all of the most common types of spatial analysis. The performance, accuracy, and effectiveness of each type of analysis were assessed using an artificial dataset with known cluster locations. The results of these analyses were then used to develop a geospatial analytical protocol to be used by scientists with little to no background in GIS or spatial analysis.

Prior to this dissertation the only types of spatial analysis that had been used to study coral diseases involved the use of various types of global measures of spatial autocorrelation (Jolles et al. 2002; Foley et al. 2005; Zvuloni et al. 2009). The global statistics used by the previous coral disease studies are designed to determine the following: (1) characterize the overall, region-wide trend of the data's spatial distribution; (2) determine what spatial scales this distribution occurs at; and (3) estimate the statistical significance associated with the spatial distribution. However, this type of *global* statistic does not identify the presence or locations of

any local variations in the spatial distribution of the data, and therefore provides no information on *where* possible disease clusters are occurring.

The geospatial analytical protocols I developed (presented in Chapter 7), expand on these global measures and incorporate the use of local measures of geospatial analysis to detect and analyze the specific locations of coral disease clusters. By applying these methods to data from a 2004 White-Band Disease (WBD) outbreak, I was able to produce mapped representations of WBD prevalence and the locations of areas with statistically significant WBD prevalence rates. Additionally, by comparing transect-level (non-weighted) and colony-level (weighted) analyses, I found that higher resolution sampling resulted in more realistic disease estimates. The results of this work, which were published earlier this year as a manuscript in *PLoS ONE* (see Lentz et al. 2011), are believed to be the first time geospatial analytical techniques have been used to visualize the spatial nature of a coral disease.

Additionally, the types of spatial analysis reviewed in this dissertation, and in particular the specific analyses that I have recommended in Chapter 7, have the potential to provide a great deal of insight into coral epizootiology. These methods can be used to study the spatio-temporal changes in coral health by comparing changes in the position, geographic distribution, and statistical significance associated with the local prevalence rates (i.e. disease clusters) for a given coral disease. By performing the same Tier 2 analyses on multiple datasets of the same coral disease from different geographic locations and/or time periods, researchers would be able to determine whether the given disease tends to have predictable geospatial distribution patterns (a sort of spatial thumbprint), or whether the spatial characteristics associated with the disease appear to vary from outbreak to outbreak. Tier 2 analyses could also be used to compare the location, size, severity, and significance of clusters associated with different coral diseases in a given location during the same period of time. The locations of disease clusters, could also be

used to guide microbial field work, such as comparing samples taken from diseased and non-diseased corals located within an area identified as having statistically significant prevalence rates to samples taken from the same types of corals located in an area with very low prevalence rates. Additionally, identifying where disease prevalence is the highest, would give marine resource managers the ability to try and protect these areas from any additional stress (such as closing that area of the reef off to tourism).

9.1.2 Availability of Spatial Coral Disease Data

The following websites have been developed over the last decade to collect global coral disease data: the Atlantic and Gulf Rapid Reef Assessment (AGRRA, *http://www.agrra.org*); the Global Coral Disease Database (GCDD; *http://coraldisease.org*, and formerly at *http://development.unep-wcmc.org/GIS/coraldis/index.cfm*); and Reefbase (*http://reefbase.org*). Of these databases, GCDD is the only one that was designed for the exclusive purpose of studying the spatial distribution of coral diseases; as is summarized by the flowing statement on their website:

> The GCDD is the result of a collaboration between UNEP-WCMC and NOAA NMFS. The project aims to collate information on the global distribution of coral diseases, in order to contribute to the understanding of coral disease prevalence. The GCDD is a compilation of information from scientific literature gathered before 2007 (archive data), as well as new contributions from users. The content of the database is being continually updated by users, creating a sustainable platform for the dissemination of coral disease data. – *http://coraldisease.org/about*

However, there are critical flaws in the design of the type of data collected by this site. First, it was designed to *only* collect coral disease data, and therefore provide no information on either the non-diseased corals or the underlying coral population at risk of the given disease. Consequently, disease *prevalence* cannot be estimated as the very definition of the word "prevalence" requires information on not only the diseased individuals but also information on

the underlying population at risk of the disease.[1] Additionally, it is impossible to assess the severity of a given disease outbreak without information on the underlying population at risk. Second, the design of the site is not conducive for performing any type of spatial analysis on their disease data, as they provide neither the coordinates of the diseased location, nor the ability to download the data. Though, prior to 2006 the GCDD did show coordinates and allow the data to be downloaded. This previous version of the database is still accessible online (*http://development.unep-wcmc.org/GIS/coraldis/index.cfm*) but is no longer being updated. However, while the previous version of the GCDD did provide the coordinates, no information was given on the scale at which the data was collected. Consequently, it is impossible to know whether the geographic coordinates of location with a given disease refers to: (A) the specific location of a diseased coral colony; (B) the location of a surveyed transect that contains one or more diseased colonies; (C) the general location of a study site in which diseased colonies were found; etc. Therefore, the results of any subsequent analysis performed on this data would be highly questionable, as there would be no way of knowing whether or not the diseased data being analyzed were referring to the same thing (i.e. data collected at the colony-level should not be combined with data that has been summarized to show diseased presence at the transect or site-level.

As with the GCDD, Reefbase also does not collect data concerning the underlying population at risk; however, the overall design of their database is far better than that of either the old or new versions of the GCDD. For example, unlike the GCDD, Reefbase's database provides the scale at which the data was collected, the source for who collected this data, and contains higher resolution location data. While both sites state that their data was collected from

[1] According to the 28th Edition of Stedman's Medical Dictionary (2006), Prevalence is defined as "the number of cases of a disease existing in a given population at a specified period of time or at a particular moment in time" (Stedman 2006, page 1559).

published literature, it appears that the coral disease data contained in both Reefbase and the old version of the GCDD was taken from the same publications. I came to this conclusion in 2006 after downloading diseased *Acropora* data from both websites and finding that while the number of diseased records was the same for both, the locations were slightly different. After closer examination, I found that this was because the coordinates provided in the GCDD has all been truncated after the second decimal place, whereas most of Reefbase's coordinates were carried out to at least four decimal places. Additionally, many of the disease records from Reefbase also contained information on the scale and original source.

Of the three online coral databases, AGRRA's is the most robust as the design of the database lacks many of the flaws listed above. For example, all of the data contained in this database were collected using AGRRA's sampling design, which facilitates analysis of data subsets within their database. Colony-level information was collected for all the corals present at the given location, and not just the corals with a particular disease. Additionally, the sampling scale and resolution of their data is not only known, but it is also clearly defined and the same for all of the records within their database. The only drawback to using the AGRRA data for spatial analysis is that the geographic coordinates provided refers to the center of each 200 by 200*m* site, rather than the location of the transects surveyed within each site. While, none of the colony-level information has been lost, by summarizing the data at the site-level (instead of the transect-level) the overall number of locations has been dramatically reduced. The reduced number of point locations limits the ability to perform different types of spatial analysis, as well as the subsequent accuracy of any spatial output. For example, their database contains information on 9,607 different coral colonies located in the Bahamas; however, any spatial analysis performed on this data would be based on the locations of sites (86 total Bahamian sites), rather than the locations of transects within each site (1004 total Bahamian transects).

Ultimately, while spatial epidemiology-related techniques have the potential to greatly improve scientific understanding of coral health, the accuracy of the results are limited by the types of data available.

9.2 Conclusions

Given the rapidly deteriorating condition of corals reefs worldwide, coupled with the grim outlook for their future, it is clear that substantial changes and progress needs to be made in the current methods being used to study coral diseases. First, there needs to be more agreement among researchers as to the nomenclature used concerning corals. Second, more detail needs to be recorded when performing disease surveys in the field. It is also important that a standardized approach and specific terminology (such as that proposed by Work and Aeby 2006, and shown in **Figure 3.9** on page 70) are used when performing these field surveys. Third, current epidemiological models need to be adapted for the marine environment, including creating alternative criteria for disease causation for cases in which Koch's postulates are not appropriate (Sutherland et al. 2004). Last, a geospatial analytical component (such as the protocols laid out in Chapter 7) needs to be added to these epidemiological models so that the spatial nature of these epizootics can be studied at local, regional, and global scales.

As the search for causation continues, surveillance and proper documentation of spatial disease patterns have the potential to not only improve scientific understanding of coral pathology, but also facilitate the conservation and protection of reefs by showing reef managers which areas of the reef are at the greatest risk. It is my hope that the material I have presented in this dissertation will provide researchers with the necessary tools and information needed for them to perform the most accurate and powerful types of geospatial analysis possible based on the data they have available, as well as assist in selecting appropriate sampling designs for future outbreak investigations.

References

Abrego D, Ulstrup KE, Willis BL, van Oppen MJH (2008) Species-specific interactions between algal endosymbionts and coral hosts define their bleaching response to heat and light stress. Proceedings of the Royal Society B-Biological Sciences 275:2273-2282

Ainsworth T, Fine M, Roff G, Hoegh-Guldberg O (2008a) Bacteria are not the primary cause of bleaching in the Mediterranean coral *Oculina patagonica*. Isme J 2:67-73

Ainsworth TD, Hoegh-Guldberg O (2008) Cellular processes of bleaching in the Mediterranean coral *Oculina patagonica*. Coral Reefs 27:593-597

Ainsworth TD, Hoegh-Guldberg O (2009) Bacterial communities closely associated with coral tissues vary under experimental and natural reef conditions and thermal stress. Aquat Biol 4:289-296

Ainsworth TD, Hoegh-Guldberg O, Leggat W (2008b) Imaging the fluorescence of marine invertebrates and their associated flora. J Microsc 232:197-199

Ainsworth TD, Kvennefors EC, Blackall LL, Fine M, Hoegh-Guldberg O (2007a) Disease and cell death in white syndrome of Acroporid corals on the Great Barrier Reef. Marine Biology 151:19-29

Ainsworth TD, Kramasky-Winter E, Loya Y, Hoegh-Guldberg O, Fine M (2007b) Coral disease diagnostics: What's between a plague and a band? Applied and Environmental Microbiology 73:981-992

Ainsworth TD, Hoegh-Guldberg O, Heron SF, Skirving WJ, Leggat W (2008c) Early cellular changes are indicators of pre-bleaching thermal stress in the coral host. Journal of Experimental Marine Biology and Ecology 364:63-71

Andrefouet S, Hochberg EJ, Chevillon C, Muller-Karger F, Brock JC, Hu C (2005) Multi-Scale Remote Sensing of Coral Reefs. In: Miller RL, Del Castillo CE, McKee BA (eds) Remote sensing of coastal aquatic environments: technologies, techniques and applications. Springer, Dordrecht, The Netherlands; Norwell, MA, pp297-315

Anselin L (1995) Local indicators of spatial association - LISA. Geogr Anal 27:93-115

Anselin L (2003) GeoDa^{TM} 0.9 User's Guide. Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois, and the Center for Spatially Integrated Social Science, Urbana, IL pp115

Anselin L (2005) Chapter 5. Spatial Statistical Modeling in a GIS Environment. In: Maguire DJ, Batty M, Goodchild MF (eds) GIS, Spatial Analysis, and Modeling. ESRI Press, Redlands, California 93-111

Anselin L, Syabri I, Kho Y (2006a) GeoDa: An introduction to Spatial Data Analysis. Geogr Anal 38:5-22

Anselin L, Lozano N, Koschinsky J (2006b) Rate Transformations and Smoothing. Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois, Urbana, IL pp85

Anthony KRN, Connolly SR, Hoegh-Guldberg O (2007) Bleaching, energetics, and coral mortality risk: Effects of temperature, light, and sediment regime. Limnology and Oceanography 52:716-726

Anthony KRN, Kline DI, Diaz-Pulido G, Dove S, Hoegh-Guldberg O (2008) Ocean acidification causes bleaching and productivity loss in coral reef builders. Proceedings of the National Academy of Sciences of the United States of America 105:17442-17446

Anthony KRN, Hoogenboom MO, Maynard JA, Grottoli AG, Middlebrook R (2009) Energetics approach to predicting mortality risk from environmental stress: a case study of coral bleaching. Funct Ecol 23:539-550

Antonius AA (1973) New observations on coral destruction in reefs. Association of Island Marine Laboratories of the Caribbean (AMLC), University of Puerto Rico (Mayaguez) 3

Antonius AA (1981) The "Band" Diseases in Coral Reefs Proceedings of the Fourth International Coral Reef Symposium, Manila, pp7-14

Aronson RB, Precht WF (1997) Stasis, biological disturbance, and community structure of a Holocene coral reef. Paleobiology 23:326-346

Aronson RB, Precht WF (2000) Herbivory and algal dynamics on the coral reef at Discovery Bay, Jamaica. Limnology and Oceanography 45:251-255

Aronson RB, Precht WF (2001) White-band disease and the changing face of Caribbean coral reefs. Hydrobiologia 460:25-38

Aronson RB, Precht WF (2006) Conservation, precaution, and Caribbean reefs. Coral Reefs 25:441-450

Aronson RB, Precht WF, Toscano MA, Koltes KH (2002a) The 1998 bleaching event and its aftermath on a coral reef in Belize. Marine Biology 141:435-447

Aronson RB, MacIntyre IG, Lewis SA, Hilbun NL (2005) Emergent zonation and geographic convergence of coral reefs. Ecology 86:2586-2600

Aronson RB, Macintyre IG, Precht WF, Murdoch TJT, Wapnick CM (2002b) The expanding scale of species turnover events on coral reefs in Belize. Ecological Monographs 72:233-249

Bailey TC, Gatrell AC (1995) Interactive Spatial Data Analysis. Longman Scientific and Technical, Essex, England pp413

Bak RPM (1983) Neoplasia, regeneration and growth in the reef-building coral *Acropora palmata*. Marine Biology 77:221-227

Baker AC (2003) Flexibility and specificity in coral-algal symbiosis: Diversity, ecology, and biogeography of *Symbiodinium*. Annu Rev Ecol Evol Syst 34:661-689

Bally M, Garrabou J (2007) Thermodependent bacterial pathogens and mass mortalities in temperate benthic communities: a new case of emerging disease linked to climate change. Global Change Biology 13:2078-2088

Balter M (1998) Infectious disease - Molecular methods fire up the hunt for emerging pathogens. Science 282:219-221

Banin E, Khare SK, Naider F, Rosenberg E (2001a) Proline-Rich Peptide from the Coral Pathogen *Vibrio shiloi* That Inhibits Photosynthesis of Zooxanthellae Applied and Environmental Microbiology 67:1536-1541

Banin E, Israely T, Fine M, Loya Y, Rosenberg E (2001b) Role of endosymbiotic zooxanthellae and coral mucus in the adhesion of the coral-bleaching pathogen *Vibrio shiloi* to its host. Fems Microbiology Letters 199:33-37

Barber RT, Hilting AK, Hayes ML (2001) The changing health of coral reefs. Human and Ecological Risk Assessment 7:1255-1270

Bascom WN, Revelle RR (1953) Free-diving: A New Exploratory Tool. American Scientist 41:624-627

Baskett ML, Nisbet RM, Kappel CV, Mumby PJ, Gaines SD (2010) Conservation management approaches to protecting the capacity for corals to respond to climate change: a theoretical comparison. Global Change Biology 16:1229-1246

Baums IB, Miller MW, Hellberg ME (2006) Geographic variation in clonal structure in a reef-building Caribbean coral, *Acropora palmata*. Ecological Monographs 76:503-519

Bayard TS, Elphick CS (2010) Using spatial point-pattern assessment to understand the social and environmental mechanisms that drive avian habitat selection. Auk 127:485-494

Beck LR, Lobitz BM, Wood BL (2000) Remote sensing and human health: New sensors and new opportunities. Emerg Infect Dis 6:217-227

Bellwood DR, Meyer CP (2009) Searching for heat in a marine biodiversity hotspot. J Biogeogr 36:569-576

Beman JM, Roberts KJ, Wegley L, Rohwer F, Francis CA (2007) Distribution and Diversity of Archaeal Ammonia Monooxygenase Genes Associated with Corals. Applied and Environmental Microbiology 73:5642-5647

Ben-Haim Y, Rosenberg E (2002) A novel *Vibrio* sp pathogen of the coral *Pocillopora damicornis*. Marine Biology 141:47-55

Ben-Haim Y, Zicherman-Keren M, Rosenberg E (2003a) Temperature-regulated bleaching and lysis of the coral *Pocillopora damicornis* by the novel pathogen *Vibrio coralliilyticus*. Applied and Environmental Microbiology 69:4236-4242

Ben-Haim Y, Thompson FL, Thompson CC, Cnockaert MC, Hoste B, Swings J, Rosenberg E (2003b) *Vibrio coralliilyticus* sp nov., a temperature-dependent pathogen of the coral *Pocillopora damicornis*. International Journal of Systematic and Evolutionary Microbiology 53:309-315

Berke O (2004) Exploratory disease mapping: kriging the spatial risk function from regional count data. International Journal of Health Geographics 3:18

Berkelmans R, Willis BL (1999) Seasonal and local spatial patterns in the upper thermal limits of corals on the inshore Central Great Barrier Reef. Coral Reefs 18:219-228

Beyer HL (2004) Hawth's Analysis Tools for ArcGIS. Available at *http://www.spatialecology.com/htools*.

Bhopal RS (2002) Concepts of epidemiology: an integrated introduction to the ideas, theories, principles, and methods of epidemiology. Oxford University Press, Oxford; New York pp317

Birkeland C (1988) Geographic comparisons of coral-reef community processes. Proceedings of the 6th International Coral Reef Symposium 1:211-220

Bithell JF (1990) An application of density estimation to geographical epidemiology. Stat Med 9:691-701

Blanton J, Manangan A, Manangan J, Hanlon C, Slate D, Rupprecht C (2006) Development of a GIS-based, real-time Internet mapping tool for rabies surveillance. International Journal of Health Geographics 5:47

Bolker BM, Altmann M, Aubert M, Ball F, Barlow ND, Bowers RG, Dobson AP, Elkington JS, Garnett GP, Gilligan CA, Hassell MP, Isham V, Jacquez JA, Kleczkowski A, Levin SA, May RM, Metz JAJ, Mollison D, Morris M, Real LA, Sattenspiel L, Swinton J, White PR, Williams BG (1995) Group Report: Spatial Dynamics of Infectious Diseases in Natural Populations. In: Grenfell BT, Dobson AP (eds) Ecology of Infectious Diseases in Natural Populations. Cambridge University Press, Cambridge, UK, pp399-420

Boulos M (2004) Descriptive review of geographic mapping of severe acute respiratory syndrome (SARS) on the Internet. International Journal of Health Geographics 3:2

Bourne D, Iida Y, Uthicke S, Smith-Keune C (2008) Changes in coral-associated microbial communities during a bleaching event. Isme J 2:350-363

Bourne DG, Munn CB (2005) Diversity of bacteria associated with the coral *Pocillopora damicornis* from the Great Barrier Reef. Environmental Microbiology 7:1162-1174

Bourne DG, Garren M, Work TM, Rosenberg E, Smith GW, Harvell CD (2009) Microbial disease and the coral holobiont. Trends Microbiol 17:554-562

Bowman A, Azzelini A (1997) Applied Smoothing Techniques for Data Analysis. Oxford University Press, Oxford pp193

Brandt ME (2009) The effect of species and colony size on the bleaching response of reef-building corals in the Florida Keys during the 2005 mass bleaching event. Coral Reefs 28:911-924

Brown BE (1997) Coral bleaching: causes and consequences. Coral Reefs 16:S129-S138

Brown BE, Dunne RP, Goodson MS, Douglas AE (2002a) Experience shapes the susceptibility of a reef coral to bleaching. Coral Reefs 21:119-126

Brown BE, Downs CA, Dunne RP, Gibb SW (2002b) Exploring the basis of thermotolerance in the reef coral *Goniastrea aspera*. Marine Ecology-Progress Series 242:119-129

Bruckner A (2002a) Proceedings of the Caribbean *Acropora* workshop: potential application of US Endangered Species Act as a conservation strategy. NOAA Technical Memorandum NMFS-OPR-24, Silver Spring, MD

Bruckner AW (2002b) Life-saving products from coral reefs. Issues Sci Technol 18:35

Bruckner AW, Bruckner RJ (1998) Rapid-wasting disease: Pathogen or predator? Science 279:2023-2025

Bruno JF, Petes LE, Harvell CD, Hettinger A (2003) Nutrient enrichment can increase the severity of coral diseases. Ecology Letters 6:1056-1061

Bruno JF, Selig ER, Casey KS, Page CA, Willis BL, Harvell CD, Sweatman H, Melendy AM (2007) Thermal stress and coral cover as drivers of coral disease outbreaks. PLoS Biol 5:1220-1227

Buddemeier RW, Kleypas JA, Aronson RB (2004) Coral Reefs and Global Climate Change – Potential Contributions of Climate change to stressors on Coral Reef Ecosystems. Pew Center on Global Climate Change 42

Burnet SM, White DO (1972) Natural History of infectious disease. Cambridge University Press, pp278

Butler RA (2005) Coral reefs decimated by 2050, Great Barrier Reef's coral 95% dead. Mongabay.com

Bythell JC, Sheppard C (1993) Mass mortality of Caribbean shallow corals. Marine Pollution Bulletin 26:296-297

Bythell JC, Barer MR, Cooney RP, Guest JR, O'Donnell AG, Pantos O, Le Tissier MDA (2002) Histopathological methods for the investigation of microbial communities associated with disease lesions in reef corals. Lett Appl Microbiol 34:359-364

Cai Q, Rushton G, Bhaduri B (2011) Validation tests of an improved kernel density estimation method for identifying disease clusters. Journal of Geographical Systems:1-22

Caldeira K, Wickett ME (2003) Anthropogenic carbon and ocean pH. Nature 425:365-365

Câmara G, Monteiro AM, Fucks SD, Carvalho MS (2008) Spatial Analaysis and GIS: A Primer. King Saud University pp30

Cao L, Caldeira K, Jain AK (2007) Effects of carbon dioxide and climate change on ocean acidification and carbonate mineral saturation. Geophysical Research Letters 34:L05607

Carilli JE, Norris RD, Black BA, Walsh SM, McField M (2009) Local Stressors Reduce Coral Resilience to Bleaching. PLoS one 4:e6324

Carilli JE, Norris RD, Black B, Walsh SM, McField M (2010) Century-scale records of coral growth rates indicate that local stressors reduce coral thermal tolerance threshold. Global Change Biology 16:1247-1257

Carlos HA, Shi X, Sargent J, Tanski S, Berke EM (2010) Density estimation and adaptive bandwidths: A primer for public health practitioners. International Journal of Health Geographics 9:39

Carlton JT, Geller JB (1993) Ecological roulette - The global transport of nonindigenous marine organisms. Science 261:78-82

Carpenter KE, Abrar M, Aeby G, Aronson RB, Banks S, Bruckner A, Chiriboga A, Cortes J, Delbeek JC, DeVantier L, Edgar GJ, Edwards AJ, Fenner D, Guzman HM, Hoeksema BW, Hodgson G, Johan O, Licuanan WY, Livingstone SR, Lovell ER, Moore JA, Obura DO, Ochavillo D, Polidoro BA, Precht WF, Quibilan MC, Reboton C, Richards ZT, Rogers AD, Sanciangco J, Sheppard A, Sheppard C, Smith J, Stuart S, Turak E, Veron JEN, Wallace C, Weil E, Wood E (2008) One-Third of Reef-Building Corals Face Elevated Extinction Risk from Climate Change and Local Impacts. Science 321:560-563

Carr AP, Rodgers AR (2002) HRE: The Home Range Extension for ArcView™ (Beta Test Version 0.9, July 1998). Tutorial Guide. Centre for Northern Forest Ecosystem Research, Ontario Ministry of Natural Resources, Ontario pp22

Carson RL (1941) Under the Sea-Wind. Oxford University Press, New York

Carson RL (1951) The Sea Around Us. Oxford University Press, USA pp288

Carson RL (1955) The Edge of the Sea. Houghton Mifflin Company, Boston

Carson RL (1962) Silent Spring. Houghton Mifflin, pp368

Casas V, Kline DI, Wegley L, Yu YN, Breitbart M, Rohwer F (2004) Widespread association of a *Rickettsiales*-like bacterium with reef-building corals. Environmental Microbiology 6:1137-1148

Causey B, Delaney J, Diaz E, Dodge D, Garcia J, Higgins J, Keller B, Kelty R, Jaap W, Matos C, Schmahl G, Rogers C, Miller M, Turgeon D (2002) Status of coral reefs in the US Caribbean and Gulf of Mexico: Florida, Texas, Puerto Rico, US Virgin Islands, Navassa. In: Wilkinson CR (ed) Status of coral reefs of the world: 2002. Australian Institute of Marine Science, Townsville 251-276

Chainey S, Ratcliffe J (2005) GIS and Crime Mapping. John Wiley & Sons, Ltd, Chichester, West Sussex, England pp428

Chalker B, Taylor D (1975) Light-enhanced calcification, and the role of oxidative phosphorylation in calcification of the coral *Acropora cervicornis*. Proceedings of the Royal Society of London Series B Biological Sciences 190:323-331

Chaput EK, Meek JI, Heimer R (2002) Spatial analysis of human granulocytic ehrlichiosis near Lyme, Connecticut. Emerg Infect Dis 8:943-948

Charlton M, Fotheringham AS (2009) Geographically Weighted Regression. National Centre for Geocomputation; National University of Ireland Maynooth., Maynooth, Co Kildare, Ireland. pp14

Charuchinda M, Hylleberg J (1984) Skeletal Extension of *Acropora formosa* at a Fringing-Reef in the Andaman Sea. Coral Reefs 3:215-219

Chimetto LA, Brocchi M, Thompson CC, Martins RCR, Ramos HR, Thompson FL (2008) Vibrios dominate as culturable nitrogen-fixing bacteria of the Brazilian coral *Mussismilia hispida*. Syst Appl Microbiol 31:312-319

Clark PF (1920) Joseph Lister, his Life and Work. The Scientific Monthly 11:518-539

Clarke KC, McLafferty SL, Tempalski BJ (1996) On epidemiology and geographic information systems: A review and discussion of future directions. Emerg Infect Dis 2:85-92

Clausen CD, Roth AA (1975) Effect of Temperature and Temperature Adaptation on Calcification Rate in Hermatypic Coral *Pocillopora damicornis*. Marine Biology 33:93-100

Cliff A (1995) Incorporating spatial components into models of epidemic spread. In: Mollison D (ed) Epidemic models: their structure and relation to data. Cambridge University Press, Cambridge, UK., pp119-149

Coffroth MA, Poland DM, Petrou EL, Brazeau DA, Holmberg JC (2010) Environmental Symbiont Acquisition May Not Be the Solution to Warming Seas for Reef-Building Corals. PLoS one 5:e13258

Coles SL, Fadlallah YH (1991) Reef coral survival and mortality at low temperatures in the Arabian Gulf: new species-specific lower temperature limits. Coral Reefs 9:231-237

Coles SL, Brown BE (2003) Coral bleaching - Capacity for acclimatization and adaptation Advances in Marine Biology, Vol 46. Academic Press Ltd, London, pp183-223

Colwell RR (1996) Global climate and infectious disease: The cholera paradigm. Science 274:2025-2031

Colwell RR, Huq A (2001) Marine ecosystems and cholera. Hydrobiologia 460:141-145

Conn HW (1895) Louis Pasteur. Science 2:601-610

Cooney RP, Pantos O, Le Tissier MDA, Barer MR, O'Donnell AG, Bythell JC (2002) Characterization of the bacterial consortium associated with black band disease in coral using molecular microbiological techniques. Environmental Microbiology 4:401-413

CoRIS (2004) Should *Acropora* spp. be Included on the Endangered Species List? NOAA's Coral Reef Information System (CoRIS), Available online at: *http://coris.noaa.gov/exchanges/acropora_esa/*

Correa AMS, Brandt ME, Smith TB, Thornhill DJ, Baker AC (2009) Symbiodinium associations with diseased and healthy scleractinian corals. Coral Reefs 28:437-448

Côté IM, Gill JA, Gardner TA, Watkinson AR (2005) Measuring coral reef decline through meta-analyses. Philosophical Transactions of the Royal Society B-Biological Sciences 360:385-395

Crabbe MJC (2008) Climate change, global warming and coral reefs: Modelling the effects of temperature. Comput Biol Chem 32:311-314

Croner CM, Sperling J, Broome FR (1996) Geographic information systems (GIS): New perspectives in understanding human health and environmental relationships. Stat Med 15:1961-1977

Croquer A, Weil E (2009) Changes in Caribbean coral disease prevalence after the 2005 bleaching event. Diseases of Aquatic Organisms 87:33-43

Crowder LB, Osherenko G, Young OR, Airame S, Norse EA, Baron N, Day JC, Bouvere F, Ehler CN, Halpern BS, Langdon SJ, McLeod KL, Ogden JC, Peach RE, Rosenberg AA, Wilson JA (2006) Sustainability—resolving mismatches in US ocean governance. Science 313:617-618

Csaszar NBM, Ralph PJ, Frankham R, Berkelmans R, van Oppen MJH (2010) Estimating the Potential for Adaptation of Corals to Climate Warming. PLoS one 5:e9751

Curtis AJ (1999) Using a spatial filter and a geographic information system to improve rabies surveillance data. Emerg Infect Dis 5:603-606

Curtis AJ, Leitner M (2006) Geographic information systems and public health: eliminating perinatal disparity. IRM Press, Hershey pp317

Curtis AJ, Lee W-AA (2010) Spatial patterns of diabetes related health problems for vulnerable populations in Los Angeles. International Journal of Health Geographics 9:1-10

Curtis AJ, Mills JW (2011) Crime in Urban Post-Disaster Environments: A Methodological Framework From New Orleans. Urban Geography 32:488-510

Curtis AJ, Duval-Diop D, Novak J (2010) Identifying spatial patterns of recovery and abandonment in the Holy Cross neighborhood of New Orleans. Cartography and Geographic Information Science 37:45-56

Dana J (1843) On the temperature limiting the distribution of corals. American Journal of Science 45:130-131

Danese M, Lazzari M, Murgante B (2008) Kernel Density Estimation Methods for a Geostatistical Appoach in Seismic Risk Analysis: The Case Study of Potenza Hilltop Town (Southern Italy). International Conference on Computational Science and Applications (ICCSA) Part 1:415-429

Daszak P, Cunningham AA, Hyatt AD (2000) Wildlife ecology - Emerging infectious diseases of wildlife - Threats to biodiversity and human health. Science 287:443-449

Daszak P, Cunningham AA, Hyatt AD (2001) Anthropogenic environmental change and the emergence of infectious diseases in wildlife. Acta Trop 78:103-116

De'ath G, Lough JM, Fabricius KE (2009) Declining Coral Calcification on the Great Barrier Reef. Science 323:116-119

de Smith MJ, Goodchild MF, Longley PA (2009) Geospatial Analysis: A comprehensive guide to principles, techniques and software tools. Matador, Leicester pp560

Denner EBM, Smith GW, Busse HJ, Schumann P, Narzt T, Polson SW, Lubitz W, Richardson LL (2003) *Aurantimonas coralicida* gen. nov., sp nov., the causative agent of white plague type II on Caribbean scleractinian corals. International Journal of Systematic and Evolutionary Microbiology 53:1115-1122

DeSalvo MK, Sunagawa S, Voolstra CR, Medina M (2010) Transcriptomic responses to heat stress and bleaching in the elkhorn coral *Acropora palmata*. Marine Ecology-Progress Series 402:97-113

Devore J, Peck R (2005) Statistics: The Exploration and Analysis of Data. Brooks/Cole -- Thomson Learning, United States pp763

Diggle P (1985) A Kernel-Method for Smoothing Point Process Data. Applied Statistics-Journal of the Royal Statistical Society Series C 34:138-147

Diggle PJ (1983) Statistical Analysis of Spatial Point Patterns. Academic Press, London, UK

Dill RF, Shumway G (1954) Geologic use of self-contained diving apparatus. Assoc Petroleum Geologists Bull 38:148-157

Doney SC, Mahowald N, Lima I, Feely RA, Mackenzie FT, Lamarque J-F, Rasch PJ (2007) Impact of anthropogenic atmospheric nitrogen and sulfur deposition on ocean acidification and the inorganic carbon system. Proceedings of the National Academy of Sciences 104:14580-14585

Donia M, Hamann MT (2003) Marine natural products and their potential applications as anti-infective agents. The Lancet Infectious Diseases 3:338-348

Donner SD, Potere D (2007) The inequity of the global threat to coral reefs. Bioscience 57:214-215

Donner SD, Knutson TR, Oppenheimer M (2007) Model-based assessment of the role of human-induced climate change in the 2005 Caribbean coral bleaching event. Proceedings of the National Academy of Sciences of the United States of America 104:5483-5488

Donner SD, Skirving WJ, Little CM, Oppenheimer M, Hoegh-Guldberg O (2005) Global assessment of coral bleaching and required rates of adaptation under climate change. Global Change Biology 11:2251-2265

Dorland WAN (1994) Dorland's Illustrated Medical Dictionary, 28th edition. W. B. Saunders Co. pp. 2000

Dorland WAN (2000) Dorland's Illustrated Medical Dictionary, 29th Edition. W.B. Saunders and Company.

Douglas AE (2003) Coral bleaching - how and why? Marine Pollution Bulletin 46:385-392

Doukan G (1957) The world beneath the waves. George Allen & Unwin

Downs CA, Woodley CM, Richmond RH, Lanning LL, Owen R (2005) Shifting the paradigm of coral-reef 'health' assessment. Marine Pollution Bulletin 51:486-494

Dubinsky Z, Falkowski P (2011) Light as a Source of Information and Energy in Zooxanthellate Corals. In: Dubinsky Z, Stambler N (eds) Coral Reefs: An Ecosystem in Transition. Springer, pp107-118

Dubinsky Z, Stambler N (2011) Coral Reefs: An Ecosystem in Transition. Springer, pp541

Dubos R (1961) Mirage of Health: Utopias, Progress and Biological. Anchor Books, Garden City, NY

Dwolatzky B, Trengrove E, Struthers H, McIntyre JA, Martinson NA (2006) Linking the global positioning system (GPS) to a personal digital assistant (PDA) to support tuberculosis control in South Africa: a pilot study. International Journal of Health Geographics 5:6

Eakin CM, Morgan JA, Heron SF, Smith TB, Liu G, Alvarez-Filip L, Baca B, Bartels E, Bastidas C, Bouchon C, Brandt M, Bruckner AW, Bunkley-Williams L, Cameron A, Causey BD, Chiappone M, Christensen TRL, Crabbe MJC, Day O, de la Guardia E, Díaz-Pulido G, DiResta D, Gil-Agudelo DL, Gilliam DS, Ginsburg RN, Gore S, Guzmán HM, Hendee JC, Hernández-Delgado EA, Husain E, Jeffrey CFG, Jones RJ, Jordán-Dahlgren E, Kaufman LS, Kline DI, Kramer PA, Lang JC, Lirman D, Mallela J, Manfrino C, Maréchal J-P, Marks K, Mihaly J, Miller WJ, Mueller EM, Muller EM, Orozco Toro CA, Oxenford HA, Ponce-Taylor D, Quinn N, Ritchie KB, Rodríguez S, Ramírez AR, Romano S, Samhouri JF, Sánchez JA, Schmahl GP, Shank BV, Skirving WJ, Steiner SCC, Villamizar E, Walsh SM, Walter C, Weil E, Williams EH, Roberson KW, Yusuf Y (2010) Caribbean Corals in Crisis: Record Thermal Stress, Bleaching, and Mortality in 2005. PLoS one 5:e13969

Earickson RJ (2007) Introduction to special issue: Eleventh International Medical Geography Symposium. Social Science & Medicine 65:1-6

Ebdon D (1985) Statistics in geography. Wiley-Blackwell, pp232

Eck JE, Chainey S, Cameron JG, Leitner M, Wilson RE (2005) Mapping Crime: Understanding Hot Spots. NIJ Special Report # 209393 pp77

Elliott P, Wakefield J, Best N, Briggs D (2006) Spatial epidemiology: methods and applications. Oxford University Press, pp475

Epstein PR, United States. National Oceanic and Atmospheric Administration. Office of Global Programs., United States. National Aeronautics and Space Agency., Health Ecological and Economic Dimensions of Global Change Program. (1998) Marine ecosystems, emerging diseases as indicators of change: health of the oceans from Labrador to Venezuela. Center for Health and the Global Environment, Harvard Medical School, Oliver Wendell Holmes Society, Boston, Mass. pp85

Erez J, Reynaud S, Silverman J, Schneider K, Allemand D (2011) Coral Calcification Under Ocean Acidification and Global Change. In: Dubinsky Z, Stambler N (eds) Coral Reefs: An Ecosystem in Transition. Springer, pp151-176

ESRI (2009a) How Spatial Autocorrelation: Moran's I (Spatial Statistics) works. ESRI, ArcGIS Desktop 9.3 Help, Available at: *http://webhelp.esri.com/ArcGISdesktop/9.3/*

ESRI (2009b) Hot Spot Analysis (Getis-Ord Gi*) (Spatial Statistics). ESRI, ArcGIS Desktop 9.3 Help, Available at: *http://webhelp.esri.com/ArcGISdesktop/9.3/*

ESRI (2009c) Cluster and Outlier Analysis: Anselin Local Moran's I (Spatial Statistics). ESRI, ArcGIS Desktop 9.3 Help, Available at: *http://webhelp.esri.com/ArcGISdesktop/9.3/*

ESRI (2009d) Geographically Weighted Regression (Spatial Statistics). ESRI, ArcGIS Desktop 9.3 Help, Available at: *http://webhelp.esri.com/ArcGISdesktop/9.3/*

ESRI (2009e) How Hot Spot Analysis: Getis-Ord Gi* (Spatial Statistics) works. ESRI, ArcGIS Desktop 9.3 Help, Available at: *http://webhelp.esri.com/ArcGISdesktop/9.3/*

ESRI (2009f) Regression analysis basics. ESRI, ArcGIS Desktop 9.3 Help, Available at: *http://webhelp.esri.com/ArcGISdesktop/9.3/*

ESRI (2009g) How Cluster and Outlier Analysis: Anselin Local Moran's I (Spatial Statistics) works. ESRI, ArcGIS Desktop 9.3 Help, Available at: *http://webhelp.esri.com/ArcGISdesktop/9.3/*

ESRI (2009h) High/Low Clustering (Getis-Ord General G) (Spatial Statistics). ESRI, ArcGIS Desktop 9.3 Help, Available at: *http://webhelp.esri.com/ArcGISdesktop/9.3/*

ESRI (2009i) Spatial Autocorrelation (Moran's I) (Spatial Statistics). ESRI, ArcGIS Desktop 9.3 Help, Available at: *http://webhelp.esri.com/ArcGISdesktop/9.3/*

ESRI (2009j) How High/Low Clustering: Getis-Ord General G (Spatial Statistics) works. ESRI, ArcGIS Desktop 9.3 Help, Available at: *http://webhelp.esri.com/ArcGISdesktop/9.3/*

Evans RD, Russ GR (2004) Larger biomass of targeted reef fish in no-take marine reserves on the Great Barrier Reef, Australia. Aquatic Conservation: Marine and Freshwater Ecosystems 14:505-519

Fabricius K (2008) Theme section on "Ocean Acidification and Coral Reefs". Coral Reefs 27:455-457

Fabricius KE (2006) Effects of irradiance, flow, and colony pigmentation on the temperature microenvironment around corals: Implications for coral bleaching? Limnology and Oceanography 51:30-37

Fager EW, Flechsig AO, Ford RF, Clutter RI, Ghelardi RJ (1966) Equipment for use in Ecological Studies using SCUBA. Limnology and Oceanography 11:503-509

Fautin DG, Romano SL (1997) Cnidaria. Sea anemones, corals, jellyfish, sea pens, hydra. Version 24 April 1997. *http://tolweb.org/Cnidaria/2461/1997.04.24* in The Tree of Life Web Project, *http://tolweb.org/*

Fautin DG, Romano SL (2000) Anthozoa. Sea Anemones, Corals, Sea Pens. Version 03 October 2000. *http://tolweb.org/Anthozoa/17634/2000.10.03* in The Tree of Life Web Project, *http://tolweb.org/*

Fautin DG, Romano SL, Jr. WAO (2000) Zoantharia. Sea Anemones and Corals. Version 04 October 2000. *http://tolweb.org/Zoantharia/17643/2000.10.04* in The Tree of Life Web Project, *http://tolweb.org/*

Feely RA, Sabine CL, Lee K, Berelson W, Kleypas J, Fabry VJ, Millero FJ (2004) Impact of anthropogenic CO_2 on the $CaCO_3$ system in the oceans. Science 305:362-366

Fine M, Tchernov D (2007a) Scleractinian coral species survive and recover from decalcification. Science 315:1811-1811

Fine M, Tchernov D (2007b) Ocean acidification and scleractinian corals - Response. Science 317:1032-1033

Fine M, Zibrowius H, Loya Y (2001) *Oculina patagonica*: a non-lessepsian scleractinian coral invading the Mediterranean Sea. Marine Biology 138:1195-1203

Fischer MM, Wang J (2011) Chapter 2: Exploring Area Data. Spatial Data Analysis. Vienna University of Economics and Business, Chinese Academy of Sciences, pp15-29

Fix E, Hodges JL (1951) Discriminatory analysis. Nonparametric discrimination; consistency properties. Report Number 4, Project Number 21-49-004. USAF School of Aviation Medicine, Randolph Field, Texas. (Reprinted as pp 261-279 of Agrawala, 1977)

Foley JE, Sokolow SH, Girvetz E, Foley CW, Foley P (2005) Spatial epidemiology of Caribbean yellow band syndrome in *Montastrea* spp. coral in the eastern Yucatan, Mexico. Hydrobiologia 548:33-40

Ford WW (1928) The Bacterial World of Joseph Lister. The Scientific Monthly 26:70-75

Fortin M-J, Dale MRT (2005) Spatial Analysis: A Guide for Ecologists. Cambridge University Press, United Kingdom pp365

Fotheringham AS (1997) Trends in quantitative methods I: Stressing the local. Prog Hum Geogr 21:88-96

Fotheringham AS (1998) Trends in quantitative methods II: stressing the computational. Prog Hum Geogr 22:283-292

Fotheringham AS, Brunsdon C, Charlton M (2000) Quantitative Geography Perspectives on Spatial Data Analysis. SAGE Publications, London, pp270

Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically Weighted Regression: the analysis of spatially varying relationships. John Wiley & Sons, West Sussex, pp269

Fredricks DN, Relman DA (1996) Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. Clin Microbiol 9:18-33

Gao S, Mio D, Anton F, Yi X, Coleman D (2008) Online GIS services for mapping and sharing disease information. International Journal of Health Geographics 7:8

Gardner TA, Cote IM, Gill JA, Grant A, Watkinson AR (2003) Long-term region-wide declines in Caribbean corals. Science 301:958-960

Garrison T (2007) Oceanography: An Invitation to Marine Science, Sixth Edition. Thomson Brooks/Cole, Canada, pp588

Garrison VH, Shinn EA, Foreman WT, Griffin DW, Holmes CW, Kellogg CA, Majewski MS, Richardson LL, Ritchie KB, Smith GW (2003) African and Asian dust: From desert soils to coral reefs. Bioscience 53:469-480

Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS (1996) Spatial point pattern analysis and its application in geographical epidemiology. Trans Inst Br Geogr NS 21:256-274

Gattuso JP, Allemand D, Frankignoulle M (1999) Photosynthesis and calcification at cellular, organismal and community levels in coral reefs: A review on interactions and control by carbonate chemistry. American Zoologist 39:160-183

Geiser DM, Taylor JW, Ritchie KB, Smith GW (1998) Cause of sea fan death in the West Indies. Nature 394:137-138

Getis A (2010) Chapter 2: Spatial Interaction and Spatial Autocorrelation: A Cross-Product Approach. In: Anselin L, Rey SJ (eds) Perspectives on Spatial Data Analysis. Springer Verlag, Berlin, pp23-33

Getis A, Ord JK (1992) The Analysis of Spatial Association by Use of Distance Statistics. Geogr Anal 24:189-206

Ginsburg RN (2000) Atlantic and Gulf Rapid Reef Assessment (AGRRA) MGG-RSMAS. University of Miami, Available online at: *http://www.agrra.org/*

Gladfelter WB (1982) White-band disease in *Acropora palmata* -- Implications for the structure and growth of shallow reefs. Bulletin of Marine Science 32:639-643

Glynn P, D'Croz L (1990) Experimental evidence for high temperature stress as the cause of El Nino-coincident coral mortality. Coral Reefs 8:181-191

Glynn PW (1993) Coral reef bleaching: ecological perspectives. Coral Reefs 12:1-17

Glynn PW, Mate JL, Baker AC, Calderon MO (2001) Coral bleaching and mortality in Panama and Ecuador during the 1997-1998 El Nino-Southern oscillation event: Spatial/temporal patterns and comparisons with the 1982-1983 event. Bulletin of Marine Science 69:79-109

Goggin JM (1960) Underwater Archaeology: Its Nature and Limitations. American Antiquity 25:348-354

Goodchild MF (1986) Spatial autocorrelation. Volume 47 in the Concepts and Techniques in Modern Geography (CATMOG) series. Geo Books, Norwich, UK, 47:56

Goreau TF, Wells JW (1967) The Shallow-Water Scleractinia of Jamaica: Revised List of Species and their Vertical Distribution Range. Bulletin of Marine Science 17:442-453

Goreau TJ, Cervino J, Goreau M, Hayes R, Hayes M, Richardson L, Smith G, DeMeyer K, Nagelkerken I, Garzon-Ferrera J, Gil D, Garrison G, Williams EH, Bunkley-Williams L, Quirolo C, Patterson K, Porter JW, Porter K (1998) Rapid spread of diseases in Caribbean coral reefs. Revista De Biologia Tropical 46:157-171

Green EP, Bruckner AW (2000) The significance of coral disease epizootiology for coral reef conservation. Biological Conservation 96:347-361

Griffith DA (2009) Spatial Autocorrelation. In: Kitchin R, Thrift N (eds) The International Encyclopedia of Human Geography. Elsevier, Oxford pp309-316

Griffith DA, Christakos G (2007) Medical geography as a science of interdisciplinary knowledge synthesis under conclusions of uncertainty. Stoch Environ Res Risk Assess 21:459-460

Grigg RW (1994) Effects of sewage discharge, fishing pressure and habitat complexity on coral ecosystems and reef fishes in Hawaii. Marine Ecology-Progress Series 103:25-34

Grimsditch GD, Salm RV (2005) Coral Reef Resilience and Resistance to Bleaching. A Global Marine Programme Working Paper. IUCN, pp50

Grober-Dunsmore R, Bonito V, Frazer TK (2006) Potential inhibitors to recovery of *Acropora palmata* populations in St. John, US Virgin Islands. Marine Ecology-Progress Series 321:123-132

Guinotte JM, Fabry VJ (2008) Ocean acidification and its potential effects on marine ecosystems. In: Ostfeld RS, Schlesinger WH (eds) Year in Ecology and Conservation Biology 2008. Blackwell Publishing, Oxford, pp320-342

Gustavson K, Huber RM, Ruitenbeek HJ (eds) (2000) Integrated Coastal Zone Management of Coral Reefs: Decision Support Modeling. World Bank Publications, Washington D.C., pp292

Haapkylä J, Unsworth RKF, Flavell M, Bourne DG, Schaffelke B, Willis BL (2011) Seasonal Rainfall and Runoff Promote Coral Disease on an Inshore Reef. PLoS one 6:e16893

Haggett P, Cliff AD, Frey AE (1977) Locational Methods in Human Geography. 2nd Edition. Edward Arnold, London, pp603

Hall P, Marron JS (1991) Lower Bounds for Bandwidth Selection in Density-Estimation. Probability Theory and Related Fields 90:149-173

Hallock P, Schlager W (1986) Nutrient excess and the demise of coral reefs and carbonate platforms. Palaios 1:389-398

Hamblin WK, Christiansen EH (2001) Earth's Dynamic Systems, 9th Edition. Prentice Hall, Upper Saddle River, New Jersey, pp735

Hardin G (1968) The Tragedy of the Commons. Science 162:1243-1248

Hardin G (1976) Carrying Capacity As an Ethical Concept. The Social Contract; Fall 2001, vol 12, Issue 1, 48-57 59:120-137

Hardy A (2003) Commentary: Bread and alum, syphilis and sunlight: rickets in the nineteenth century. International Journal of Epidemiology 32:337-340

Harley CDG, Hughes AR, Hultgren KM, Miner BG, Sorte CJB, Thornber CS, Rodriguez LF, Tomanek L, Williams SL (2006) The impacts of climate change in coastal marine systems. Ecology Letters 9:228-241

Harrison PL (2011) Sexual Reproduction of Scleractinian Corals. In: Dubinsky Z, Stambler N (eds) Coral Reefs: An Ecosystem in Transition. Springer, pp59-85

Harvell CD (2004) Ecology and evolution of host-pathogen interactions in nature. American Naturalist 164:S1-S5

Harvell CD, Kim K, Quirolo C, Weir J, Smith G (2001) Coral bleaching and disease: contributors to 1998 mass mortality in *Briareum asbestinum* (Octocorallia, Gorgonacea). Hydrobiologia 460:97-104

Harvell CD, Mitchell CE, Ward JR, Altizer S, Dobson AP, Ostfeld RS, Samuel MD (2002) Ecology - Climate warming and disease risks for terrestrial and marine biota. Science 296:2158-2162

Harvell CD, Jordan-Dahlgren E, Merkel S, Rosenberg E, Raymundo L, Smith G, Weil E, Willis B (2007) Coral Disease, Environmental Drivers, and the Balance Between Coral and Microbial Associates. Oceanography 20:172-195

Harvell CD, Kim K, Burkholder JM, Colwell RR, Epstein PR, Grimes DJ, Hofmann EE, Lipp EK, Osterhaus A, Overstreet RM, Porter JW, Smith GW, Vasta GR (1999) Review: Marine ecology - Emerging marine diseases - Climate links and anthropogenic factors. Science 285:1505-1510

Harvell CD, Aronson R, Baron N, Connell J, Dobson A, Ellner S, Gerber L, Kim K, Kuris A, McCallum H, Lafferty K, McKay B, Porter J, Pascual M, Smith G, Sutherland K, Ward J (2004) The rising tide of ocean diseases: unsolved problems and research priorities. Frontiers in Ecology and the Environment 2:375-382

Harvell D, Altizer S, Cattadori IM, Harrington L, Weil E (2009) Climate change and wildlife diseases: When does the host matter the most? Ecology 90:912-920

Hayes ML, Bonaventura J, Mitchell TP, Prospero JM, Shinn EA, Van Dolah F, Barber RT (2001) How are climate and marine biological outbreaks functionally linked? Hydrobiologia 460:213-220

Hayes RL, Goreau NI (1998) The significance of emerging diseases in the tropical coral reef ecosystem. Rev Biol Trop 46 Supl. 5:173-185

Hazelton M (1996) Bandwidth selection for local density estimators. Scandinavian Journal of Statistics 23:221-232

Highsmith RC (1982) Reproduction by fragmentation in corals. Marine ecology progress series 7:207-226

Hoegh-Guldberg O (1999) Climate change, coral bleaching and the future of the world's coral reefs. Marine and Freshwater Research 50:839-866

Hoegh-Guldberg O (2004) Coral reefs in a century of rapid environmental change. Symbiosis 37:1-31

Hoegh-Guldberg O (2005) Low coral cover in a high-CO_2 world. Journal of Geophysical Research-Oceans 110:C09S06

Hoegh-Guldberg O (2011) The Impact of Climate Change on Coral Reef Ecosystems. In: Dubinsky Z, Stambler N (eds) Coral Reefs: An Ecosystem in Transition. Springer, pp391-403

Hoegh-Guldberg O, Fine M (2004) Low temperatures cause coral bleaching. Coral Reefs 23:444-444

Hoegh-Guldberg O, Fine M, Skirving W, Johnstone R, Dove S, Strong A (2005) Coral bleaching following wintry weather. Limnology and Oceanography 50:265-271

Hoegh-Guldberg O, Mumby PJ, Hooten AJ, Steneck RS, Greenfield P, Gomez E, Harvell CD, Sale PF, Edwards AJ, Caldeira K, Knowlton N, Eakin CM, Iglesias-Prieto R, Muthiga N, Bradbury RH, Dubi A, Hatziolos ME (2007) Coral reefs under rapid climate change and ocean acidification. Science 318:1737-1742

Hoerling MP, Hurrell JW, Xu TY (2001) Tropical origins for recent North Atlantic climate change. Science 292:90-92

Holmes E (1997) Basic Epidemiological concepts in a spatial context. In: Tilman D, Kareiva PD (eds) Spatial Ecology. Princeton University Press, Princeton, New Jersey, pp111-136

Hu W, Clements A, Williams G, Tong S (2011) Spatial analysis of notified dengue fever infections. Epidemiol Infect 139:391-399

Hubbard D (1997) Reefs as Dynamic Systems. In: Birkeland C (ed) Life and Death of Coral Reefs. International Thomson Publishing, New York, pp43-67

Hubbard DK, Gladfelter EH, Bythell JC (1994) Comparison of biological and geological perspectives of coral-reef community structure at Buck Island, U.S. Virgin Islands. Colloquium on Global Aspects of Coral Reefs: Health, Hazards and History

Hubbard DK, Burke RB, Gill IP, Ramirez WR, Sherman C (2008) Ch. 7 Coral-reef Geology: Puerto Rico and the US Virgin Islands. In: Riegl BM, Dodge RE (eds) Coral Reefs of the USA. Springer, pp263-302

Hughes TP (1994) Catastrophes, phase-shifts, and large-scale degradation of a Caribbean coral reef. Science 265:1547-1551

Humann P, Deloach N (2002) Reef Coral Identification: Florida, Caribbean, Bahamas, Including Marine Plants. New World Publications, Inc, Jacksonville, FL, pp287

ICRI/UNEP-WCMC (2010a) Disease in Tropical Coral Reef Ecosystems: ICRI Key Messages on Coral Disease. pp11

ICRI/UNEP-WCMC (2010b) ICRI Good practice guidance on coral disease monitoring. pp13

Jaap WC (1979) Observations on zooxanthellae expulsion at Middle Sambo reef, Florida Keys. Bulletin of Marine Science 29:414-422

Jackson JBC (1997) Reefs since Columbus. Coral Reefs 16:S23-S32

Jackson JBC (2001) What was natural in the coastal oceans? Proceedings of the National Academy of Sciences of the United States of America 98:5411-5418

Jackson JBC, Kirby MX, Berger WH, Bjorndal KA, Botsford LW, Bourque BJ, Bradbury RH, Cooke R, Erlandson J, Estes JA, Hughes TP, Kidwell S, Lange CB, Lenihan HS, Pandolfi JM, Peterson CH, Steneck RS, Tegner MJ, Warner RR (2001) Historical overfishing and the recent collapse of coastal ecosystems. Science 293:629-638

James NP (1974) Diagenesis of scleractinian corals in the subaerial vadose environment. Journal of Paleontology:785-799

Johnson PTJ, Townsend AR, Cleveland CC, Glibert PM, Howarth RW, McKenzie VJ, Rejmankova E, Ward MH (2010) Linking environmental nutrient enrichment and disease emergence in humans and wildlife. Ecological Applications 20:16-29

Johnston I, Rohwer F (2007) Microbial landscapes on the outer tissue surfaces of the reef-building coral *Porites compressa*. Coral Reefs 26:375-383

Jokiel PL (2004) Chapter 23. Temperature Stress and Coral Bleaching. In: Rosenberg E, Loya Y (eds) Coral Health and Disease. Springer, New York, pp401-425

Jokiel PL, Coles SL (1977) Effects of Temperature on Mortality and Growth of Hawaiian Reef Corals. Marine Biology 43:201-208

Jolles AE, Sullivan P, Alker AP, Harvell CD (2002) Disease transmission of aspergillosis in sea fans: Inferring process from spatial pattern. Ecology 83:2373-2378

Jones AM, Berkelmans R, van Oppen MJH, Mieog JC, Sinclair W (2008) A community change in the algal endosymbionts of a scleractinian coral following a natural bleaching event: field evidence of acclimatization. Proceedings of the Royal Society B-Biological Sciences 275:1359-1365

Jones MC, Marron JS, Sheather SJ (1996) A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association 91:401-407

Jones SD (2004) Mapping a zoonotic disease: Anglo-American efforts to control bovine tuberculosis before World War I. Osiris 19:133-148

Joseph AE (1985) Measuring potential access to health services: some comments on the design and calibration of indices. First International Medical Geography Symposium, Nottingham, England

Kaczmarsky LT, Draud M, Williams EH (2005) Is there a relationship between proximity to sewage effluent and the prevalence of coral disease. Caribbean Journal of Science 41:124-137

Kellogg CA (2004) Tropical Archaea: diversity associated with the surface microlayer of corals. Marine Ecology Progress Series 273:81-88

Kellogg WW (1987) Mankind's impact on climate: The evolution of an awareness. Climatic Change 10:113-136

Kelman D, Kashman Y, Rosenberg E, Kushmaro A, Loya Y (2006) Antimicrobial activity of Red Sea corals. Marine Biology 149:357-363

Kidson C, Steers J, Flemming N (1962) A trial of the potential value of aqualung diving to coastal physiography on British coasts. The Geographical journal 128:49-53

Kim K, Harvell CD (2002) Aspergillosis of sea fan corals: disease dynamics in the Florida Keys. In: Porter JW, Porter KG (eds) The Everglades, Florida Bay, and Coral Reefs of the Florida Keys: An Ecosystem Sourcebook CRC Press, Boca Raton, pp813-824

Kim K, Kim PD, Alker AP, Harvell CD (2000a) Chemical resistance of gorgonian corals against fungal infections. Marine Biology 137:393-401

Kim K, Harvell CD, Kim PD, Smith GW, Merkel SM (2000b) Fungal disease resistance of Caribbean sea fan corals (*Gorgonia* spp.). Marine Biology 136:259-267

Kinlan BP, Gaines SD (2003) Propagule dispersal in marine and terrestrial environments: A community perspective. Ecology 84:2007-2020

Kinsman DJJ (1964) Reef Coral Tolerance of High Temperatures and Salinities. Nature 202:1280-1282

Klaus JS, Frias-Lopez J, Bonheyo GT, Heikoop JM, Fouke BW (2005) Bacterial communities inhabiting the healthy tissues of two Caribbean reef corals: interspecific and spatial variation. Coral Reefs 24:129-137

Klaus JS, Janse I, Heikoop JM, Sanford RA, Fouke BW (2007) Coral microbial communities, zooxanthellae and mucus along gradients of seawater depth and coastal pollution. Environmental Microbiology 9:1291-1305

Kleppel G, Dodge R, Reese C (1989) Changes in pigmentation associated with the bleaching of stony corals. Limnology and Oceanography 34:1331-1335

Kleypas JA, McManus JW, Menez LAB (1999a) Environmental limits to coral reef development: Where do we draw the line? American Zoologist 39:146-159

Kleypas JA, Buddemeier RW, Archer D, Gattuso JP, Langdon C, Opdyke BN (1999b) Geochemical consequences of increased atmospheric carbon dioxide on coral reefs. Science 284:118-120

Kline DI, Kuntz NM, Breitbart M, Knowlton N, Rohwer F (2006) Role of elevated organic carbon levels and microbial activity in coral mortality. Marine Ecology-Progress Series 314:119-125

Koch T (2005) Cartographies of disease: maps, mapping, and medicine. ESRI Press, Redlands, CA, pp389

Koch T (2008) John Snow, hero of cholera: RIP. Canadian Medical Association journal 178:1736-1736

Koch T, Denike K (2009) Crediting his critics' concerns: Remaking John Snow's map of Broad Street cholera, 1854. Social Science & Medicine 69:1246-1251

Koren O, Rosenberg E (2006) Bacteria associated with mucus and tissues of the coral *Oculina patagonica* in summer and winter. Applied and Environmental Microbiology 72:5254-5259

Korrubel JL, Riegl B (1998) A new coral disease from the southern Arabian Gulf. Coral Reefs 17:22-22

Krediet CJ, Ritchie KB, Cohen M, Lipp EK, Sutherland KP, Teplitski M (2009) Utilization of Mucus from the Coral *Acropora palmata* by the Pathogen *Serratia marcescens* and by Environmental and Coral Commensal Bacteria. Applied and Environmental Microbiology 75:3851-3858

Kulldorff M (2006) SaTScanTM (v. 7.0), Available at: *http://www.satscan.org/*

Kulldorff M (2010) SaTScan^{TM} User Guide for version 9.0. pp109

Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F (2005) A space-time permutation scan statistic for disease outbreak detection. PLos Med 2:216-224

Kumpf HE, Randall HA (1961) Charting the Marine Environments of St. John, U.S. Virgin Islands. Bulletin of Marine Science 11:543-551

Kushmaro A, Loya Y, Fine M, Rosenberg E (1996) Bacterial infection and coral bleaching. Nature 380:396-396

Kushmaro A, Rosenberg E, Fine M, Loya Y (1997) Bleaching of the coral *Oculina patagonica* by *Vibrio* AK-1. Marine Ecology-Progress Series 147:159-165

Kushmaro A, Rosenberg E, Fine M, Ben Haim Y, Loya Y (1998) Effect of temperature on bleaching of the coral *Oculina patagonica* by *Vibrio* AK-1. Marine Ecology-Progress Series 171:131-137

Kushmaro A, Banin E, Loya Y, Stackebrandt E, Rosenberg E (2001) *Vibrio shiloi* sp nov., the causative agent of bleaching of the coral *Oculina patagonica*. International Journal of Systematic and Evolutionary Microbiology 51:1383-1388

Kvennefors ECE, Sampayo E, Kerr C, Vieira G, Roff G, Barnes AC (2011) Regulation of Bacterial Communities Through Antimicrobial Activity by the Coral Holobiont. Microb Ecol 63:605-618

Lancaster J, Downes BJ (2004) Spatial point pattern analysis of available and exploited resources. Ecography 27:94-102

Lang JC (2003) Status of Coral Reefs in the Western Atlantic: Results of Initial Surveys, Atlantic and Gulf Rapid Reef Assessment (AGGRA) Program. Atoll Research Bulletin 496:0-630

Lang JC, Lasker HR, Gladfelter EH, Hallock P, Jaap WC, Losada FJ, Muller RG (1992) Spatial and Temporal Variability During Periods of Recovery After Mass Bleaching on Western Atlantic Coral Reefs. American Zoologist 32:696-706

Lawson A (2006) Statistical methods in spatial epidemiology. Wiley, Chichester, pp398

Lee J, Marion LK (1994) Analysis of Spatial Autocorrelation of U.S.G.S. 1:250,000 Digital Elevation Models. GIS/LIS:504-513

Legendre P (1993) Spatial Autocorrelation -- Trouble or New Paradigm. Ecology 74:1659-1673

Leggat W, Hoegh-Guldberg O, Dove S, Yellowlees D (2007a) Analysis of an EST library from the dinoflagellate (*Symbiodinium* sp.) symbiont of reef-building corals. J Phycol 43:1010-1021

Leggat W, Ainsworth T, Bythell J, Dove S, Gates R, Hoegh-Guldberg O, Iglesias-Prieto R, Yellowlees D (2007b) The hologenome theory disregards the coral holobiont. Nat Rev Microbiol 5(10)

Leichter JJ, Stewart HL, Miller SL (2003) Episodic nutrient transport to Florida coral reefs. Limnology and Oceanography 48:1394-1407

Leitner M, Brecht H (2007) Crime analysis and mapping with GeoDa 0.9-5-i. Soc Sci Comput Rev 25:265-271

Lentz JA (2005) Home Range and Habitat Preferences of *Terrapene carolina carolina* at Jug Bay Wetlands Sanctuary, Maryland. Undergraduate Thesis Ph.D. thesis, Hamilton College, pp340

Lentz JA, Blackburn JK, Curtis AJ (2011) Evaluating Patterns of a White-Band Disease (WBD) Outbreak in *Acropora palmata* Using Spatial Analysis: A Comparison of Transect and Colony Clustering. PLoS one 6:e21830

Leopold A (1966) A Sand County Almanac. Oxford University Press, New York, pp269

Lesser MP (2004) Experimental biology of coral reef ecosystems. J Exp Biol Ecol 300:217-252

Lesser MP (2007) Coral reef bleaching and global climate change: Can corals survive the next century? Proceedings of the National Academy of Sciences of the United States of America 104:5259-5260

Lesser MP (2011) Coral Bleaching: Causes and Mechanisms. In: Dubinsky Z, Stambler N (eds) Coral Reefs: An Ecosystem in Transition. Springer, pp405-419

Lesser MP, Bythell JC, Gates RD, Johnstone RW, Hoegh-Guldberg O (2007) Are infectious diseases really killing corals? Alternative interpretations of the experimental and ecological data. Journal of Experimental Marine Biology and Ecology 346:36-44

Levine N (2007) *CrimeStat.* A Spatial Statistics Program for the Analysis of Crime Inci-dent Locations (v 3.1). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. March.

Levine N, Associates (2004) CrimeStat® III, version 3.0: A Spatial Statistics Program for the Analysis of Crime Incident Locations. The National Institute of Justice, Washington, D.C. pp939

Levine N, Associates (2009) CrimeStat® III, version 3.2a Update Notes. The National Institute of Justice, Washington, D.C. pp97

Lewis JB (1984) The *Acropora* Inheritance: A Reinterpretation of the Development of Fringing Reefs in Barbados, West Indies. Coral Reefs 3:117-122

Link MC (1959) Sea diver: a quest for history under the sea. Rinehart

Lins-de-Barros M, Vieira R, Cardoso A, Monteiro V, Turque A, Silveira C, Albano R, Clementino M, Martins O (2010) Archaea, Bacteria, and Algal Plastids Associated with the Reef-Building Corals *Siderastrea stellata* and *Mussismilia hispida* from Búzios, South Atlantic Ocean, Brazil. Microb Ecol 59:523-532

Lirman D (1999) Reef fish communities associated with *Acropora palmata*: Relationships to benthic attributes. Bulletin of Marine Science 65:235-252

Lirman D (2000) Fragmentation in the branching coral *Acropora palmata* (Lamarck): growth, survivorship, and reproduction of colonies and fragments. Journal of Experimental Marine Biology and Ecology 251:41-57

Lirman D, Schopmeyer S, Manzello D, Gramer LJ, Precht WF, Muller-Karger F, Banks K, Barnes B, Bartels E, Bourque A, Byrne J, Donahue S, Duquesnel J, Fisher L, Gilliam D, Hendee J, Johnson M, Maxwell K, McDevitt E, Monty J, Rueda D, Ruzicka R, Thanner S (2011) Severe 2010 Cold-Water Event Caused Unprecedented Mortality to Corals of the Florida Reef Tract and Reversed Previous Survivorship Patterns. PLoS one 6:e23047

Lo CP, Yeung AKW (2007) Concepts and Techniques of Geographic Information Systems. Second Edition. Pearson Prentice Hall, Upper Saddle River, NJ, pp532

Longley PA, Goodchild MF, Maguire DJ, Rhind DW (2005) Geographic Information Systems and Science. John Wiley & Sons, Ltd, Barcelona, pp517

Lunine JI (1999) Earth -- Evolution of a Habitable World. Cambridge University Press, Cambridge, UK, pp319

Lutz PE (1986) Invertebrate Zoology. The Benjamin/Cummings Publishing Company, Inc., Menlo Park, California, pp734

Macintyre IG (1967) Submerged Coral Reefs, West Coast of Barbados, West Indies. Can J Earth Sci 4:461-474

Madl P (2005) The Silent Sentinels: The Demise of Tropical Coral Reefs. University of Salzburg, Electronic Textbook, Available online at: *http://biophysics.sbg.ac.at/reefs/reefs.htm*

Maguire DJ, Batty M, Goodchild MF (eds) (2005) GIS, Spatial Analysis, and Modeling. ESRI Press, Redlands, California, pp480

Maher MM (2010) Lining up data in ArcGIS: a guide to map projections. ESRI Press, Redlands, California, pp184

Maina J, McClanahan TR, Venus V, Ateweberhan M, Madin J (2011) Global Gradients of Coral Exposure to Environmental Stresses and Implications for Local Management. PLoS one 6:e23064

Marcogliese DJ (2001) Implications of climate change for parasitism of animals in the aquatic environment. Can J Zool-Rev Can Zool 79:1331-1352

Marcon E, Puech F (2009) Generalizing Ripley's K function to inhomogeneous populations. HAL - CCSD, halshs-00372631: 21

Marshall AT, Clode P (2004) Calcification rate and the effect of temperature in a zooxanthellate and an azooxanthellate scleractinian reef coral. Coral Reefs 23:218-224

Marshall P, Schuttenberg H (2006) A Reef Manager's Guide to Coral Bleaching. Great Barrier Reef Marine Park Authority, Townsville, MC pp178

Marshall PA, Baird AH (2000) Bleaching of corals on the Great Barrier Reef: differential susceptibilities among taxa. Coral Reefs 19:155-163

Mayor AG (1914) The effects of temperature upon tropical marine animals. Carnegie Institution of Washington 183:1-24

Mayor AG (1915) The lower temperature at which reef-corals lose their ability to capture food. Carnegie Inst Yearbook 14:212

Mayor PA, Rogers CS, Hillis-Starr ZM (2006) Distribution and abundance of elkhorn coral, *Acropora palmata*, and prevalence of White-Band disease at Buck Island Reef National Monument, St. Croix, US Virgin Islands. Coral Reefs 25:239-242

McCallum H, Barlow N, Hone J (2001) How should pathogen transmission be modelled? Trends in Ecology & Evolution 16:295-300

McCallum H, Harvell D, Dobson A (2003) Rates of spread of marine pathogens. Ecology Letters 6:1062-1067

McCallum HI, Kuris A, Harvell CD, Lafferty KD, Smith GW, Porter J (2004) Does terrestrial epidemiology apply to marine systems? Trends in Ecology & Evolution 19:585-591

McClanahan TR, Muthiga NA (1998) An ecological shift in a remote coral atoll of Belize over 25 years. Environmental Conservation 25:122-130

McClanahan TR, Muthiga NA, Mangi S (2001) Coral and algal changes after the 1998 coral bleaching: interaction with reef management and herbivores on Kenyan reefs. Coral Reefs 19:380-391

McClanahan TR, Baird AH, Marshall PA, Toscano MA (2004a) Comparing bleaching and mortality responses of hard corals between southern Kenya and the Great Barrier Reef, Australia. Marine Pollution Bulletin 48:327-335

McClanahan TR, McLaughlin SM, Davy JE, Wilson WH, Peters EC, Price KL, Maina J (2004b) Observations of a new source of coral mortality along the Kenyan coast. Hydrobiologia 530-31:469-479

McCreedy C, Miller J, Charles CW, Rogers CS (2006) Response to coral bleaching in U.S. Virgin Islands National Park. In: National Park Service U (ed) Water Resources Division. U.S. department of the Interior pp

McGill SA (2000) Louis Pasteur. History Remembers Scientists of the Past 3181515:15

McLeod KS (2000) Our sense of Snow: the myth of John Snow in medical geography. Social Science & Medicine 50:923-935

Meade MS, Earickson RJ (2000) Medical geography. Guilford Press, New York City

Meade MS, Earickson R (2005) Medical geography. Guilford Press, New York, pp501

MedicineNet (2004) Definition of Medical Geography. MedicineNet, Available online at: *http://www.medterms.com*

MedicineNet (2011) MedTerms Medical Dictionary. MedicineNet.com, Available online at: *http://www.medterms.com*

MedlinePlus (2003) MedlinePlus (online) in conjunction with Merriam Webster Online Dictionary. Available online at: *http://www.nlm.nih.gov/medlineplus/mplusdictionary.html*

Mescher K, Sturgess A (2009) Coral Reef Recovery from Hurricane Damage and Implications of Coral Reefs for Future Medical Discoveries. Epistimi:27-30

Middlebrook R, Hoegh-Guldberg O, Leggat W (2008) The effect of thermal history on the susceptibility of reef-building corals to thermal stress. J Exp Biol 211:1050-1056

Milius S (1998) Bacteria Cause Plague in Coral Reef. Science News 153:229

Miller J, Muller E, Rogers C, Waara R, Atkinson A, Whelan KRT, Patterson M, Witcher B (2009) Coral disease following massive bleaching in 2005 causes 60% decline in coral cover on reefs in the US Virgin Islands. Coral Reefs 28:925-937

Mills JW, Curtis A, Pine JC, Kennedy B, Jones F, Ramani R, Bausch D (2008) The clearinghouse concept: a model for geospatial data centralization and dissemination in a disaster. Disasters 32:467-479

Miloslavich P, Díaz JM, Klein E, Alvarado JJ, Díaz C, Gobin J, Escobar-Briones E, Cruz-Motta JJ, Weil E, Cortés J, Bastidas AC, Robertson R, Zapata F, Martín A, Castillo J, Kazandjian A, Ortiz M (2010) Marine Biodiversity in the Caribbean: Regional Estimates and Distribution Patterns. PLoS one 5:e11916

Morrison AC, Astete H, Chapilliquen F, Ramirez-Prada G, Diaz G, Getis A, Gray K, Scott TW (2004) Evaluation of a sampling methodology for rapid assessment of *Aedes aegypti* infestation levels in Iquitos, Peru. Journal of Medical Entomology 41:502-510

Mullen K, Harvell CD, Peters E (2004) Coral Resistance to Disease. In: Rosenberg E, Loya Y (eds) Coral Health and Disease. Springer, pp377-399

Mydlarz LD, McGinty ES, Harvell CD (2010) What are the physiological and immunological responses of coral to climate warming and disease? J Exp Biol 213:934-945

Myint SW (2010) Spatial Autocorrelation. In: Warf B (ed) Encyclopedia of Geography. Sage Publications Inc., London and Thousand Oaks, CA pp2607-2608

Nagelkerken I, Buchan K, Smith GW, Bonair K, Bush P, Garzon-Ferreira J, Botero L, Gayle P, Harvell CD, Heberer C, Kim K, Petrovic C, Pors L, Yoshioka P (1997) Widespread disease in Caribbean sea fans: II. Patterns of infection and tissue loss. Marine Ecology-Progress Series 160:255-263

Neumann AC, Macintyre I (1985) Reef Response to Sea Level Rise: Keep-up, Catch-up or Give-up Proceedings of the Fifth International Coral Reef Congress, Tahiti 3:105-110

Nixon SW (1995) Coastal marine eutrophication -- A definition, social causes, and future concerns. Ophelia 41:199-219

NMFS (2006) Endangered and Threatened Species: Final Listing Determinations for Elkhorn Coral and Staghorn Coral. National Marine Fisheries Service (NMFS), Federal Register, 50 CFR Part 223 [Docket No. 050304058-6116-03; I.D. No., 050204C], RIN No. 054B-XB29, pp26852-26861

Nowak R (2004) Sewage nutrients fuel coral disease. New Scientist 181:12-13

O'Shea ML, Field R (1992) Detection and Disinfection of Pathogens in Storm-Generated Flows. Canadian Journal of Microbiology 38:267-276

OBIS (2011) Ocean Biogeographic Information System (OBIS): A Data-Sharing Legacy of the Census of Marine Life. Available online at: http://www.iobis.org/.

Ohde S, Hossain MMM (2004) Effect of $CaCO_3$ (aragonite) saturation state of seawater on calcification of *Porites* coral. Geochem J 38:613-621

Openshaw S (1984) The Modifiable Areal Unit Problem. Concepts and Techniques in Modern Geography. Geo Abstracts Ltd, Norwich, England pp41

Openshaw S, Taylor PJ (1979) A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem. In: Wrigley N (ed) Statistical Applications in the Spatial Sciences. Pion, London, pp127-144

Ormsby T, Napoleon E, Burke R, Groessl C, Feaster L (2001) Getting to know ArcGIS desktop: basics of ArcView, ArcEditor, and ArcInfo. Second Edition updated for ArcGIS 9. ESRI Press, Redlands, California, pp572

Orr JC, Fabry VJ, Aumont O, Bopp L, Doney SC, Feely RA, Gnanadesikan A, Gruber N, Ishida A, Joos F, Key RM, Lindsay K, Maier-Reimer E, Matear R, Monfray P, Mouchet A, Najjar RG, Plattner GK, Rodgers KB, Sabine CL, Sarmiento JL, Schlitzer R, Slater RD, Totterdell IJ, Weirig MF, Yamanaka Y, Yool A (2005) Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. Nature 437:681-686

Ostfeld RS, Glass GE, Keesing F (2005) Spatial epidemiology: an emerging (or re-emerging) discipline. Trends in Ecology & Evolution 20:328-336

Page C, Willis B (2006) Distribution, host range and large-scale spatial variability in black band disease prevalence on the Great Barrier Reef, Australia. Diseases of Aquatic Organisms 69:41-51

Pandolfi JM (2011) The Paleoecology of Coral Reefs. In: Dubinsky Z, Stambler N (eds) Coral Reefs: An Ecosystem in Transition. Springer, pp13-24

Pandolfi JM, Jackson JBC, Baron N, Bradbury RH, Guzman HM, Hughes TP, Kappel CV, Micheli F, Ogden JC, Possingham HP, Sala E (2005) Ecology - Are US coral reefs on the slippery slope to slime? Science 307:1725-1726

Pandolfi JM, Bradbury RH, Sala E, Hughes TP, Bjorndal KA, Cooke RG, McArdle D, McClenachan L, Newman MJH, Paredes G, Warner RR, Jackson JBC (2003) Global trajectories of the long-term decline of coral reef ecosystems. Science 301:955-958

Pantos O, Bythell JC (2006) Bacterial community structure associated with white band disease in the elkhorn coral *Acropora palmata* determined using culture-independent 16S rRNA techniques. Diseases of Aquatic Organisms 69:79-88

Patterson KL, Porter JW, Ritchie KE, Polson SW, Mueller E, Peters EC, Santavy DL, Smiths GW (2002) The etiology of white pox, a lethal disease of the Caribbean elkhorn coral, *Acropora palmata*. Proceedings of the National Academy of Sciences of the United States of America 99:8725-8730

Perry GLW, Miller BP, Enright NJ (2006) A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. Plant Ecol 187:59-82

Peters EC (1984) A survey of cellular reactions to environmental stress and disease in Caribbean scleractinian corals. Helgoland Mar Res 37:113-137

Peters EC (1997) Chapter 6. Diseases of coral reef organisms. In: Birkeland C (ed) Life and Death of Coral Reefs. Chapman & Hall, New York, pp114-139

Peters EC (2006) Diseases of Corals and Other Reef Organisms. Graduate-level course held at Mote Marine Laboratory's Tropical Research Laboratory, Summerland Key, FL. July 29-August 6, 2006.

Peters EC, Oprandy JJ, Yevich PP (1983) Possible causal agent of white band disease in Caribbean Acroporid corals. J Invertebr Pathol 41:394-396

Pittman SJ, Brown KA (2011) Multi-Scale Approach for Predicting Fish Species Distributions across Coral Reef Seascapes. PLoS one 6:e20583

Plummer M (2003) Movement and Home Range Ecology. Ecology Laboratory, Department of Biology, Harding University, Available online at: *http://www.harding.edu/USER/plummer/WWW/ecology/ecolab/movementlab.htm*

Pohl O (2003) Dust clouds may carry infectious organisms across oceans. Scientific American 289:18-19

Porter JW, Meier OW (1992) Quantification of loss and change in Floridian reef coral populations. American Zoologist 32:625-640

Porter JW, Dustan P, Jaap WC, Patterson KL, Kosmynin V, Meier OW, Patterson ME, Parsons M (2001) Patterns of spread of coral disease in the Florida Keys. Hydrobiologia 460:1-24

Power AG, Mitchell CE (2004) Pathogen spillover in disease epidemics. American Naturalist 164:S79-S89

Precht WF, Robbart ML, Aronson RB (2004) The potential listing of *Acropora* species under the US Endangered Species Act. Marine Pollution Bulletin 49:534-536

Precht WF, Bruckner AW, Aronson RB, Bruckner RJ (2002) Endangered acroporid corals of the Caribbean. Coral Reefs 21:41-42

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) Numerical recipes: the art of scientific computing. Cambridge Univ. Press, Cambridge, UK, pp818

Ratcliffe JH (2010) The spatial dependency of crime increase dispersion. Secur J 23:18-36

Raymundo LJ (2005) Coral Disease as an Emerging Management Issue International Coral Reef Initiative (ICRI), Koror, Palau pp4

Raymundo LJ, Couch CS, Harvell CD (eds) (2008) Coral Disease Handbook: Guidelines for Assessment, Monitoring and Management. Coral Reef Targeted Research and Capacity Building for Management Program. University of Queensland, Centre for Marine Studies, St Lucia QLD, Australia. pp121

Real LA (1996) Disease ecology. Ecology 77:989-989

Real LA, McElhany P (1996) Spatial pattern and process in plant-pathogen interactions. Ecology 77:1011-1025

Real LA, Marschall EM, Roche BM (1992) Individual behavior and pollination ecology: implications for sexually transmitted plant diseases. In: DeAngelis DL, Gross LJ (eds) Individual-based models and approaches in ecology. Chapman and Hall, New York, USA., pp492-508

Reef R, Kaniewska P, Hoegh-Guldberg O (2009) Coral Skeletons Defend Against Ultraviolet Radiation. PLoS one 4(11):e7995

Reis A, Araújo Jr S, Moura R, Francini-Filho R, Pappas Jr G, Coelho A, Krüger R, Thompson F (2009) Bacterial diversity associated with the Brazilian endemic reef coral *Mussismilia braziliensis*. J Appl Microbiol 106:1378-1387

Reshef L, Koren O, Loya Y, Zilber-Rosenberg I, Rosenberg E (2006) The Coral Probiotic Hypothesis. Environmental Microbiology 8:2068-2073

Revelle R, Suess HE (1957) Carbon Dioxide Exchange Between Atmosphere and Ocean and the Question of an Increase of Atmospheric CO_2 during the Past Decades. Tellus 9:18-27

Revelle R, Broecker W, Craig H, Keeling CD, Smagorinsky J (1965) Restoring the Quality of our Environment: Report of the Environmental Pollution Panel, President's Science Advisory Committee. The White House, Washington, DC pp. 111-133

Rezaeian M, Dunn G, St Leger S, Appleby L (2007) Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. J Epidemiol Community Health 61:98-102

Richardson D (1999) A brief history of scuba diving in the United States. SPUMS 29(3):173-176

Richardson LL (1998) Coral diseases: what is really known? Trends in Ecology & Evolution 13:438-443

Richardson LL, Goldberg WM, Carlton RG, Halas JC (1998a) Coral disease outbreak in the Florida Keys: Plague Type II. Revista De Biologia Tropical 46:187-198

Richardson LL, Smith GW, Ritchie KB, Carlton RG (2001) Integrating microbiological, microsensor, molecular, and physiologic techniques in the study of coral disease pathogenesis. Hydrobiologia 460:71-89

Richardson LL, Goldberg WM, Kuta KG, Aronson RB, Smith GW, Ritchie KB, Halas JC, Feingold JS, Miller SL (1998b) Florida's mystery coral-killer identified. Nature 392:557-558

Richmond RH (1993) Coral reefs -- Present problems and future concerns resulting from anthropogenic disturbance. American Zoologist 33:524-536

Richmond RH, Wolanski E (2011) Coral Research: Past Efforts and Future Horizons. In: Dubinsky Z, Stambler N (eds) Coral Reefs: An Ecosystem in Transition. Springer, pp3-10

Riegl B (2002) Effects of the 1996 and 1998 positive sea-surface temperature anomalies on corals, coral diseases and fish in the Arabian Gulf (Dubai, UAE). Marine Biology 140:29-40

Riegl B, Bruckner A, Coles SL, Renaud P, Dodge RE (2009) Coral Reefs Threats and Conservation in an Era of Global Change Year in Ecology and Conservation Biology 2009. Blackwell Publishing, Oxford, pp136-186

Ritchie KB (2006) Regulation of microbial populations by coral surface mucus and mucus-associated bacteria. Marine Ecology-Progress Series 322:1-14

Ritchie KB, Smith GW (1998) Type II white-band disease. Revista De Biologia Tropical 46:199-203

Ritchie KB, Polson SW, Smith GW (2001) Microbial disease causation in marine invertebrates: problems, practices, and future prospects. Hydrobiologia 460:131-139

Roberts H, Rouse L, Walker ND, Hudson J (1982) Cold-water stress in Florida Bay and northern Bahamas; a product of winter cold-air outbreaks. Journal of Sedimentary Research 52:145-155

Roberts JA (1968) SCUBA - Tool for Coastal Geographic Research. Prof Geogr 20:420-422

Robertson C, Nelson TA (2010) Review of software for space-time disease surveillance. International Journal of Health Geographics 9:8

Rodgers AR, Carr AP (1998) HRE: The Home Range Extension for ArcView. Ontario Ministry of Natural Resources, Centre for Northern Forest Ecosystem Research, Thunder Bay, Ontario, Canada. Available online at: *http://flash.lakeheadu.ca/~arodgers/hre/*.

Rodgers AR, Carr AP (2002) HRE: The Home Range Extension for ArcView™ (Beta Test Version 0.9, July 1998). User's Manual. Centre for Northern Forest Ecosystem Research, Ontario Ministry of Natural Resources, Ontario pp27

Rodgers AR, Kie JG (2010) HRT: Home Range Tools for ArcGIS(R) (Version 1.1, June 2007). User's Manual (Draft September 28, 2010). Centre for Northern Forest Ecosystem Research, Ontario Ministry of Natural Resources, Ontario pp26

Rogers CS (1990) Responses of coral reefs and reef organisms to sedimentation. Marine Ecology-Progress Series 62:185-202

Rohwer F, Seguritan V, Azam F, Knowlton N (2002) Diversity and distribution of coral-associated bacteria. Mar Ecol-Prog Ser 243:1-10

Rohwer FR, Breitbart MB, Jara JJ, Azam FA, Knowlton NK (2001) Diversity of bacteria associated with the Caribbean coral *Montastraea franksi*. Coral Reefs 20:85-91

Romano SL, Cairns SD (2002) Romano, Sandra L. and Stephen D. Cairns. 2002. Scleractinia. Stony star corals. Version 28 October 2002 (under construction). *http://tolweb.org/Scleractinia/17653/2002.10.28* in The Tree of Life Web Project, *http://tolweb.org/*

Rose A (2009) Coral Taxonomy. *http://www.coralscience.org/articles/PDF/Coral%20taxonomy.pdf*, in Coral Publications, *http://www.coralscience.org*, Chamber of Commerce, Utrecht.

Rosenberg E (2004) Chapter 25. The Bacterial Disease Hypothesis of Coral Bleaching. In: Rosenberg E, Loya Y (eds) Coral Health and Disease. Springer, New York, pp445-461

Rosenberg E, Ben-Haim Y (2002) Microbial diseases of corals and global warming. Environmental Microbiology 4:318-326

Rosenberg E, Falkovitz L (2004) The *Vibrio shiloi*/*Oculina patagonica* model system of coral bleaching. Annu Rev Microbiol 58:143-159

Rosenberg E, Barash Y (2005) Chapter 18. Microbial Diseases of Corals. In: Belkin S, Colwell RR (eds) Oceans and Health: Pathogens in the Marine Environment. Springer, New York, pp415-430

Rosenberg E, Zilber-Rosenberg I (2011) Symbiosis and Development: The Hologenome Concept. Birth Defects Res Part C-Embryo Today-Rev 93:56-66

Rosenberg E, Kushmaro A (2011) Microbial Diseases of Corals: Pathology and Ecology. In: Dubinsky Z, Stambler N (eds) Coral Reefs: An Ecosystem in Transition. Springer, pp451-464

Rosenberg E, Kellogg CA, Rohwer F (2007a) Coral Microbiology. Oceanography 20:146-154

Rosenberg E, Sharon G, Zilber-Rosenberg I (2009a) The hologenome theory of evolution contains Lamarckian aspects within a Darwinian framework. Environmental Microbiology 11:2959-2962

Rosenberg E, Sharon G, Atad I, Zilber-Rosenberg I (2010) The evolution of animals and plants via symbiosis with microorganisms. Environ Microbiol Rep 2:500-506

Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I (2007b) The role of microorganisms in coral health, disease and evolution. Nat Rev Micro 5:355-362

Rosenberg E, Kushmaro A, Kramarsky-Winter E, Banin E, Yossi L (2009b) The role of microorganisms in coral bleaching. Isme J 3:139-146

Rosenshein L, Scott L, Pratt M (2011) Finding a Meaningful Model: This checklist will help you evaluate regression models. ArcUser. ESRI, Redlands, CA pp40-45

Ruiz-Moreno D, Pascual M, Emch M, Yunus M (2010) Spatial clustering in the spatio-temporal dynamics of endemic cholera. BMC Infect Dis 10:51

Rushton G, Lolonis P (1996) Exploratory spatial analysis of birth defect rates in an urban population. Stat Med 15:717-726

Rushton G, Krishnamurthy R, Krishnamurti D, Lolonis P, Song H (1996) The spatial relationship between infant mortality and birth defect rates in a US city. Stat Med 15:1907-1919

Rushton G, Peleg I, Banerjee A, Smith G, West MM (2004) Analyzing Geographic Patterns of Disease Incidence: Rates of Late-Stage Colorectal Cancer in Iowa. Journal of Medical Systems 28:224-236

Rypien KL, Ward JR, Azam F (2010) Antagonistic interactions among coral-associated bacteria. Environmental Microbiology 12:28-39

Sain SR, Baggerly KA, Scott DW (1994) Cross-Validation of multivariate densities. Journal of the American Statistical Association 89:807-817

Sammarco PW (1982) Polyp Bail-Out: An Escape Response to Environmental Stress and a New Means of Reproduction in Corals. Marine Ecology-Progress Series 10:57-65

SCEP (1970) Man's impact on the global environment: assessment and recommendations for action. Study of Critical Environmental Problems (SCEP), Massachusetts Institute of Technology (M.I.T.P.), Cambridge, MA, pp318

Schmid WD (1965) Distribution of Aquatic Vegetation as Measured by Line Intercept with SCUBA. Ecology 46:816-823

Schmidt G, Wolfe J (2009) Climate Change: Picturing the Science. W. W. Norton & Company, New York, pp305

Schutte VGW, Selig ER, Bruno JF (2010) Regional spatio-temporal trends in Caribbean coral reef benthic communities. Marine Ecology-Progress Series 402:115-122

Segar DA, Segar ES (2007) Introduction to Ocean Sciences, Second Edition. W. W. Norton & Company, New York, pp581

Selig ER, Casey KS, Bruno JF (2010) New insights into global patterns of ocean temperature anomalies: implications for coral reef health and management. Glob Ecol Biogeogr 19:397-411

Selig ER, Harvell CD, Bruno JF, Willis BL, Page CA, Casey KS, Sweatman H (2006) Analyzing the Relationship Between Ocean Temperature Anomalies and Coral Disease Outbreaks at Broad Spatial Scales. In: Phinney J, Hoegh-Guldberg O, Kleypas J, Skirving W, Strong A (eds) Coral reefs and climate change: Science and management American Geophysical Union, Washington (D. C.), pp111 - 128

Selkoe KA, Halpern BS, Ebert CM, Franklin EC, Selig ER, Casey KS, Bruno J, Toonen RJ (2009) A map of human impacts to a "pristine" coral reef ecosystem, the PapahAnaumokuAkea Marine National Monument. Coral Reefs 28:635-650

Shashar N, Banaszak AT, Lesser MP, Amrami D (1997) Coral Endolithic Algae: Life in a Protected Environment. Pacific Science 51:167-173

Sheppard CRC, Davy SK, Pilling GM (2009) The Biology of Coral Reefs. Oxford University Press, New York, pp339

Shinn E (1963) Spur and groove formation on the Florida reef tract. Journal of Sedimentary Research 33:291-303

Shinn EA (1976) Coral reef recovery in Florida and the Persian Gulf. Environ Geol 1:241-254

Shinn EA (2004) The mixed value of environmental regulations: do acroporid corals deserve endangered species status? Marine Pollution Bulletin 49:531-533

Shinn EA, Reese RS, Reich CD (1994) Fate and pathways of injection-well effluent in the Florida Keys. US Geological Survey, Open-File Report # 94-276

Shinn EA, Smith GW, Prospero JM, Betzer P, Hayes ML, Garrison V, Barber RT (2000) African dust and the demise of Caribbean coral reefs. Geophysical Research Letters 27:3029-3032

Siboni N, Ben-Dov E, Sivan A, Kushmaro A (2008) Global distribution and diversity of coral-associated Archaea and their possible role in the coral holobiont nitrogen cycle. . Environ Microbiol 10:2979-2990

Siegenthaler U, Stocker TF, Monnin E, Luthi D, Schwander J, Stauffer B, Raynaud D, Barnola JM, Fischer H, Masson-Delmotte V, Jouzel J (2005) Stable carbon cycle-climate relationship during the late Pleistocene. Science 310:1313-1317

Sigl W, Vonrad U, Oeltzsch.H, Braune K, Fabriciu.F (1969) Diving Sled - A Tool to Increase Efficiency of Underwater Mapping by SCUBA Divers. Marine Geology 7:357

Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London, pp175

Silverman BW, Jones MC (1989) E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951). International Statistical Review / Revue Internationale de Statistique 57:233-238

Silverman J, Lazar B, Cao L, Caldeira K, Erez J (2009) Coral reefs may start dissolving when atmospheric CO_2 doubles. Geophysical Research Letters 36:L05606

Singleton P, Sainsbury D (2006) Dictionary of Microbiology and Molecular Biology, Third Edition. Wiley, Chichester, West Sussex; Hoboken, NJ pp895

Smith GW, Harvell CD, Kim K (1998) Response of sea fans to infection with *Aspergillus* sp. (Fungi). Rev Biol Trop 46:205-208

Smith GW, Ives LD, Nagelkerken IA, Ritchie KB (1996) Caribbean sea-fan mortalities. Nature 383:487-487

Smith SC, Bruce CW (2008) CrimeStat III User Workbook The National Institute of Justice, Washington, DC pp117

Snow SJ (2002) Commentary: Sutherland, Snow and water: the transmission of cholera in the nineteenth century. International Journal of Epidemiology 31:908-911

Snow SJ (2008) The art of medicine - John Snow: the making of a hero? Lancet 372:22-23

Snyder MA (1966) Exploring Underwater with SCUBA. Arch Environ Health 13(6):690-694

Sokolow S (2009) Effects of a changing climate on the dynamics of coral infectious disease: a review of the evidence. Diseases of Aquatic Organisms 87:5-18

Sokolow SH, Foley P, Foley JE, Hastings A, Richardson LL (2009) Editor's choice: Disease dynamics in marine metapopulations: modelling infectious diseases on coral reefs. J Appl Ecol 46:621-631

Spalding M, Bunting G (2004) A guide to the coral reefs of the Caribbean. University of California Press, pp256

Spalding M, Ravilious C, Green EP (2001) World Atlas of Coral Reefs. University of California Press, Berkley, CA, pp.424

Spatz C, Johnston JO (1976) Basic Statistics: Tales of Distributions. Brooks/Cole Publishing Company, Monterey, California, pp380

Squires DF (1965) Neoplasia in a coral. Science 148:503

Stat M, Loh WKW, LaJeunesse TC, Hoegh-Guldberg O, Carter DA (2009) Stability of coral-endosymbiont associations during and after a thermal stress event in the southern Great Barrier Reef. Coral Reefs 28:709-713

Stedman TL (2006) Stedman's Medical Dictionary. Lippincott Williams & Wilkins, Philadelphia, pp2162

Strychar KB, Coates M, Sammarco PW (2004a) Loss of *Symbiodinium* from bleached Australian scleractinian corals (*Acropora hyacinthus*, *Favites complanata* and *Porites solida*). Marine and Freshwater Research 55:135-144

Strychar KB, Coates M, Sammarco PW, Piva TJ (2004b) Bleaching as a pathogenic response in scleractinian corals, evidenced by high concentrations of apoptotic and necrotic zooxanthellae. Journal of Experimental Marine Biology and Ecology 304:99-121

Sumich JL, Morrissey JF (2004) Introduction to the Biology of Marine Life, Eighth Edition. Jones and Bartlett Publishers, Sudbury, Massachusetts, pp449

Sunagawa S, DeSalvo MK, Voolstra CR, Reyes-Bermudez A, Medina M (2009) Identification and Gene Expression Analysis of a Taxonomically Restricted Cysteine-Rich Protein Family in Reef-Building Corals. PLoS one 4(3):e4865

Sutherland KP, Porter JW, Torres C (2004) Disease and immunity in Caribbean and Indo-Pacific zooxanthellate corals. Marine Ecology-Progress Series 266:273-302

Sutherland KP, Shaban S, Joyner JL, Porter JW, Lipp EK (2011) Human Pathogen Shown to Cause Disease in the Threatened Eklhorn Coral *Acropora palmata*. PLoS one 6:e23468

Sutherland KP, Porter JW, Turner JW, Thomas BJ, Looney EE, Luna TP, Meyers MK, Futch JC, Lipp EK (2010) Human sewage identified as likely source of white pox disease of the threatened Caribbean elkhorn coral, *Acropora palmata*. Environmental Microbiology 12:1122-1131

Szmant AM (1986) Reproductive ecology of Caribbean reef corals. Coral Reefs 5:43-53

Szmant AM (2002) Nutrient enrichment on coral reefs: Is it a major cause of coral reef decline? Estuaries 25:743-766

Szmant AM, Gassman NJ (1990) The effects of prolonged “bleaching” on the tissue biomass and reproduction of the reef coral *Montastrea annularis*. Coral Reefs 8:217-224

Tabangin DR, Flores JC, Emperador NF (2010) Investigating Crime Hotspot Places and their Implication to Urban Environmental Design: A Geographic Visualization and Data Mining Approach. International Journal of Human and Social Sciences 5(4):210-218

Takahashi S, Whitney S, Itoh S, Maruyama T, Badger M (2008) Heat stress causes inhibition of the *de novo* synthesis of antenna proteins and photobleaching in cultured *Symbiodinium*. Proceedings of the National Academy of Sciences of the United States of America 105:4203-4208

Talbot TO, Kulldorff M, Forand SP, Haley VB (2000) Evaluation of spatial filters to create smoothed maps of health data. Stat Med 19:2399-2408

Terrell GR (1990) The Maximal Smoothing Principle in Density Estimation. Journal of the American Statistical Association 85:470-477

Thompson FL, Barash Y, Sawabe T, Sharon G, Swings J, Rosenberg E (2006) *Thalassomonas loyana* sp nov, a causative agent of the white plague-like disease of corals on the Eilat coral reef. International Journal of Systematic and Evolutionary Microbiology 56:365-368

Thomson DA (1957) Investigations of the Applications of SCUBA and Free Swimming Technics in Fresh and Salt Water Fishery Biology., University of Michigan, pp44

Thurber RV, Willner-Hall D, Rodriguez-Mueller B, Desnues C, Edwards RA, Angly F, Dinsdale E, Kelly L, Rohwer F (2009) Metagenomic analysis of stressed coral holobionts. Environmental Microbiology 11:2148-2163

Tiwari C, Rushton G (2005) Using Spatially Adaptive Filters to Map Late Stage Colorectal Cancer Incidence in Iowa. In: Fisher P (ed) Developments in Spatial Data Handling. Springer Berlin Heidelberg, pp665-676

Toledo-Hernández C, Zuluaga-Montero A, Bones-González A, Rodríguez J, Sabat A, Bayman P (2008) Fungi in healthy and diseased sea fans (*Gorgonia ventalina*): is *Aspergillus sydowii* always the pathogen? Coral Reefs 27:707-714

Trevelyan B, Smallman-Raynor M, Cliff AD (2005) The spatial dynamics of poliomyelitis in the United States: From epidemic emergence to vaccine-induced retreat, 1910-1971. Ann Assoc Am Geogr 95:269-293

Trujillo AP, Thurman HV (2008) Essentials of Oceanography (9th Edition). Pearson Prentice Hall, Upper Saddle River, NJ, pp534

Tu J, Xia ZG (2008) Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. Sci Total Environ 407:358-378

US EPA (2000) Stressor identification document. Office of Water, Office of Research and Development, United States Environmental Protection Agency (US EPA), Washington, DC

Van Gelder RN (2002) Koch's postulates and the polymerase chain reaction. Ocul Immunol Inflamm 10:235-238

van Oppen MJH, McDonald BJ, Willis B, Miller DJ (2001) The evolutionary history of the coral genus *Acropora* (Scleractinia, Cnidaria) based on a mitochondrial and a nuclear marker: Reticulation, incomplete lineage sorting, or morphological convergence? Molecular Biology and Evolution 18:1315-1329

Vandenbroucke JP (2001) Changing images of John Snow in the history of epidemiology. Sozial-Und Praventivmedizin 46:288-293

Vandermeulen JH, Davis ND, Muscatine L (1972) The effect of inhibitors of photosynthesis on zooxanthellae in corals and other marine invertebrates. Marine Biology 16:185-191

Vaughan TW (1918) The temperature of the Florida coral-reef tract. Carnegie Institute of Washington Publication 213:321-339

Vaughan TW (1919) Corals and the formation of coral reefs. Annual Report of the Smithsonian Institute 17:189-238

Vaughan TW, Wells JW (1943) Revision of the suborders, families, and genera of the Scleractinia. Geological Society of America Special Papers 44:55-56

Veron JEN (2000) Corals of the World. Australian Institute of Marine Science, Townsville, Australia

Veron JEN (2008) Mass extinctions and ocean acidification: biological constraints on geological dilemmas. Coral Reefs 27:459-472

Veron JEN, Wallace CC (1984) Scleractinia of Eastern Australia. Part I. Family Acroporidae. Australian Institute of Marine Science Monograph Series vol. 6, pp1385

Veron JEN, Hoegh-Guldberg O, Lenton TM, Lough JM, Obura DO, Pearce-Kelly P, Sheppard CRC, Spalding M, Stafford-Smith MG, Rogers AD (2009) The coral reef crisis: The critical importance of < 350 ppm CO_2. Marine Pollution Bulletin 58:1428-1436

Vitousek PM (1994) Beyond global warming -- Ecology and global change. Ecology 75:1861-1876

Vollmer SV, Kline DI (2008) Natural Disease Resistance in Threatened Staghorn Corals. PLoS ONE 3:e3718

Voolstra CR, Sunagawa S, Matz MV, Bayer T, Aranda M, Buschiazzo E, DeSalvo MK, Lindquist E, Szmant AM, Coffroth MA, Medina M (2011) Rapid Evolution of Coral Proteins Responsible for Interaction with the Environment. PLoS one 6:e20392

Wade T, Sommer S (2006) A to Z GIS: an illustrated dictionary of geographic information systems. ESRI Press: Independent Publishers Group [distributor], Redlands, Calif., pp268

Walker ND, Roberts HH, Rouse LJ, Huh OK (1982) Thermal History of Reef-Associated Environments During a Record Cold-Air Outbreak Event. Coral Reefs 1:83-87

Wallace CC (1999) Staghorn corals of the world : a revision of the coral genus *Acropora* (Scleractinia; Astrocoeniina; Acroporidae) worldwide, with emphasis on morphology, phylogeny and biogeography. CSIRO Pub., Collingwood, Vic., pp421

Waller LA, Gotway CA (2004) Applied spatial statistics for public health data. John Wiley & Sons, Hoboken, N.J., pp520

Walton Smith FG (1971) Atlantic Reef Corals. A Handbook of the Common Reef and Shallow-Water Corals of Bermuda, The Bahamas, Florida, The West Indies and Brazil. University of Miami Press, Coral Gables, Florida, pp164

Wand MP, Jones MC (1995) Kernel smoothing. Chapman and Hall, Ltd., London, UK, pp212

Wang F (2006) Quantitative methods and applications in GIS. CRC/Taylor & Francis, Boca Raton, FL, pp265

Wapnick CM, Precht WF, Aronson RB (2004) Millennial-scale dynamics of staghorn coral in Discovery Bay, Jamaica. Ecology Letters 7:354-361

Ward JR, Kim K, Harvell CD (2007) Temperature affects coral disease resistance and pathogen growth. Marine Ecology-Progress Series 329:115-121

Watkins PJ (2011) A family of doctors over 250 years: innovation and controversy. Journal of Medical Biography 19:56-62

Wegley L, Yu Y, Breitbart M, Casas V, Kline DI, Rohwer F (2004) Coral-associated Archaea. Marine Ecology Progress Series 273:89-96

Weil E (2004) Chapter 2. Coral reef diseases in the wider Caribbean. In: Rosenberg E, Loya Y (eds) Coral Health and Disease. Springer, New York, pp35-68

Weil E, Croquer A (2009) Spatial variability in distribution and prevalence of Caribbean scleractinian coral and octocoral diseases. I. Community-level analysis. Diseases of Aquatic Organisms 83:195-208

Weil E, Rogers CS (2011) Coral Reef Diseases in the Atlantic-Caribbean. In: Dubinsky Z, Stambler N (eds) Coral Reefs: An Ecosystem in Transition. Springer, pp465-491

Weil E, Urreiztieta I, Garzón-Ferreira J (2002) Geographic variability in the incidence of coral and octocoral diseases in the wider Caribbean. Proc 9th Int Coral Reef Symp 2:1231-1237

Western D (2001) Human-modified ecosystems and future evolution. PNAS 98:5458-5465

Whittaker RH (1959) On the Broad Classification of Organisms. Q Rev Biol 34:210-226

Whittaker RH (1969) New Concepts of Kingdoms of Organisms. Science 163:150-160

Whittaker RH, Margulis L (1978) Protist Classification and Kingdoms of Organisms. Biosystems 10:3-18

Wiegand T, Moloney KA (2004) Rings, circles, and null-models for point pattern analysis in ecology. Oikos 104:209-229

Wiegand T, Gunatilleke S, Gunatilleke N, Okuda T (2007) Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. Ecology 88:3088-3102

Williams DE, Miller MW (2005) Coral disease outbreak: pattern, prevalence and transmission in *Acropora cervicornis*. Marine Ecology-Progress Series 301:119-128

Williams DE, Miller MW, Kramer KL (2008) Recruitment failure in Florida Keys *Acropora palmata*, a threatened Caribbean coral. Coral Reefs 27:697-705

Williams GJ, Aeby GS, Cowie ROM, Davy SK (2010) Predictive Modeling of Coral Disease Distribution within a Reef System. PLoS one 5:e9264

Williamson D, McLafferty S, Goldsmith V, McGuire P, Mollenkopf J (1998) Smoothing Crime Incident Data: New Methods for Determining the Bandwidth in Kernel Estimation ESRI User Conference

Williamson D, McLafferty S, Goldsmith V, Mollenkopf J, McGuire P (1999) A Better Method To Smooth Crime Incident Data ESRI ArcUser Magazine pp1-5

Willis BL, Page CA, Dinsdale EA (2004) Chapter 3. Coral disease on the Great Barrier Reef. . In: Rosenberg E, Loya Y (eds) Coral Health and Disease. Springer-Verlag, Berlin., pp69-104

Wilson SK, Adjeroud M, Bellwood DR, Berumen ML, Booth D, Bozec Y-M, Chabanet P, Cheal A, Cinner J, Depczynski M, Feary DA, Gagliano M, Graham NAJ, Halford AR, Halpern BS, Harborne AR, Hoey AS, Holbrook SJ, Jones GP, Kulbiki M, Letourneur Y, De Loma TL, McClanahan T, McCormick MI, Meekan MG, Mumby PJ, Munday PL, Ohman MC, Pratchett MS, Riegl B, Sano M, Schmitt RJ, Syms C (2010) Crucial knowledge gaps in current understanding of climate change impacts on coral reef fishes. J Exp Biol 213:894-900

Wilson WH, Dale AL, Davy JE, Davy SK (2005) An enemy within? Observations of virus-like particles in reef corals. Coral Reefs 24:145-148

Winkler R, Antonius A, Renegar DA (2004) The skeleton eroding band disease on coral reefs of Aqaba, Red Sea. Marine Ecology-Pubblicazioni Della Stazione Zoologica Di Napoli I 25:129-144

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences 87:4576-4579

Wood R (1998) The ecological evolution of reefs. Annual Review of Ecology and Systematics 29:179-206

Wood RD (1963) Adapting SCUBA to Aquatic Plant Ecology. Ecology 44:416-419

Woodley CM, Bruckner AW, McLenon AL, Higgins JL, B. GS, Nicholson JH (2008) Field Manual for Investigating Coral Disease Outbreaks. NOAA Technical Memorandum NOS NCCOS 80 and CRCP 6. National Oceanic and Atmospheric Administration (NOAA), Silver Spring, MD. pp85

Wooldridge SA, Done TJ (2009) Improved water quality can ameliorate effects of climate change on corals. Ecological Applications 19:1492-1499

Work TM, Aeby GS (2006) Systematically describing gross lesions in corals. Diseases of Aquatic Organisms 70:155-160

Work TM, Richardson LL, Reynolds TL, Willis BL (2008) Biomedical and veterinary science can increase our understanding of coral disease. Journal of Experimental Marine Biology and Ecology 362:63-70

Worton BJ (1989) Kernel methods for estimating the utilization distribution in home-range studies. Ecology 70:164-168

Worton BJ (1995) A convex hull-based estimator of home-range size. Biometrics 51:1206-1215

Zilber-Rosenberg I, Rosenberg E (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. Fems Microbiol Rev 32:723-735

Zubillaga AL, Marquez LM, Croquer A, Bastidas C (2008) Ecological and genetic data indicate recovery of the endangered coral *Acropora palmata* in Los Roques, Southern Caribbean. Coral Reefs 27:63-72

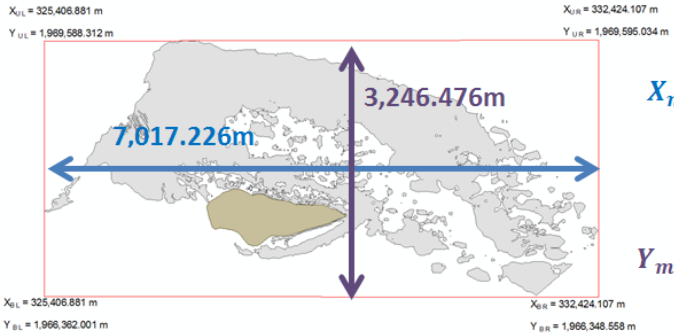
Zvuloni A, Artzy-Randrup Y, Stone L, Kramarsky-Winter E, Barkan R, Loya Y (2009) Spatio-Temporal Transmission Patterns of Black-Band Disease in a Coral Community. PLoS one 4:e4993

Appendix A – Spatial Parameter Estimation

A.1 Grid Cell Resolution Estimation Methods

"We suggest following the methodology by Ratcliffe (1999) where cell size resolution is the result of dividing the **shorter side** of the **minimum bounding rectangle** (i.e. the shortest of the two extents between the maximum x and minimum x, and the maximum y and minimum y) by 150." [1] p. 159

*Grid Cell Resolution Estimated using **UTM coordinates***



Horizontal Distance

$$X_{max} - X_{min} = 332{,}424.107m - 325{,}406.881m$$
$$= \mathbf{7{,}017.226m}$$

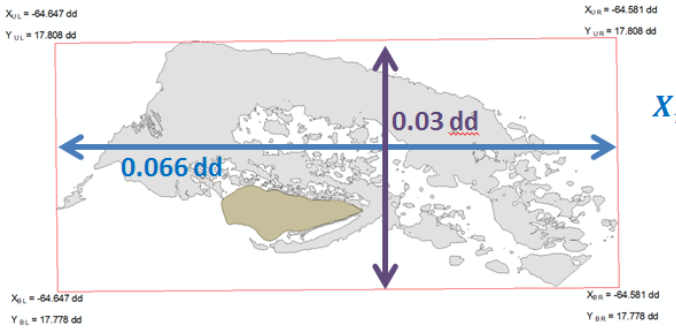
Verticle Distance

$$Y_{max} - Y_{min} = 1{,}969{,}595.034 - 1{,}966{,}348.558m$$
$$= \mathbf{3{,}246.476m}$$

The shortest side is the Vertical Distance, so the $\boldsymbol{grid\ cell\ size} = \frac{3246.476m}{150} = \mathbf{21.643173\overline{3}\ m}$

OR

*Grid Cell Resolution Estimated using **UTM coordinates***



Horizontal Distance

$$X_{max} - X_{min} = -64.581dd - (-64.647dd)$$
$$= \mathbf{0.066\ dd}$$

Vertical Distance

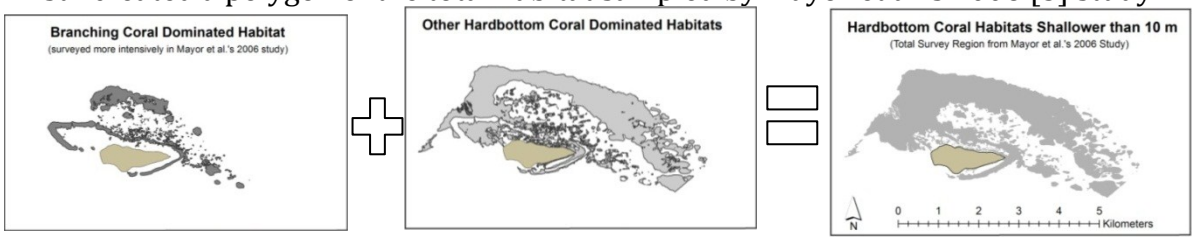
$$Y_{max} - Y_{min} = 17.808dd - 17.778dd$$
$$= \mathbf{0.03\ dd}$$

The shortest side is the vertical distance, so the $\boldsymbol{grid\ cell\ size} = \frac{0.03\ dd}{150} = \mathbf{0.0002}\ decimal\ degrees$

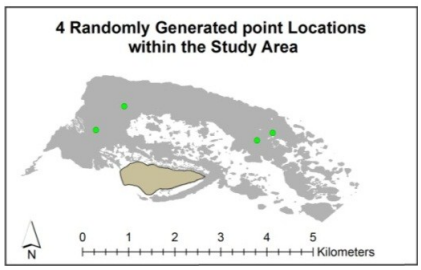
A.2 Visual Calibration using an Artificial Cluster Dataset

Creating the Artificial Cluster Dataset [2]

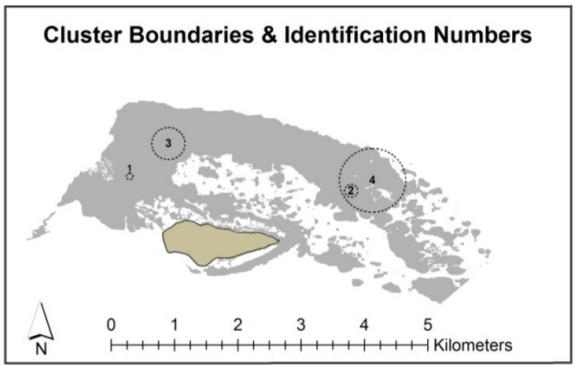
1) First I created a polygon of the total habitat sampled by Mayor et al.'s 2006 [3] study:



2) *Hawth's Tools extension* was used to generate 4 random point locations within this polygon

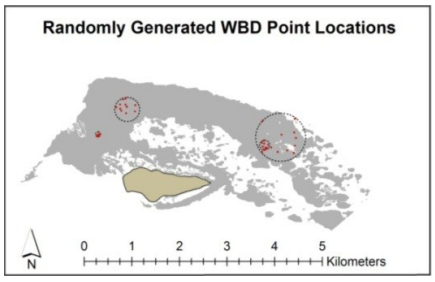


3) I used these 4 points to define the centers of 4 clusters.
   - Clusters were by generating a circle with a certain radius size around each point.
   - Radii of 50*m*, 100*m*, 250*m*, & 500*m*, were chosen in order to test the accuracy of the spatial analysis software on detecting clusters of different sizes within the same dataset.
   - The radii were assigned to the clusters based on the associated Cluster ID number , resulting in the following clusters – radii combinations: Cluster 1—50*m*, Cluster 2—100*m*, Cluster 3—250*m*, & Cluster 4—500*m*.
   - The boundaries of these newly defined clusters, along with the identification number of each cluster were then mapped.
   - This resulted in Cluster 2 falling within the boundaries of Cluster 4. I decided to keep this, in order to see how the spatial analysis software handled the situation [4].



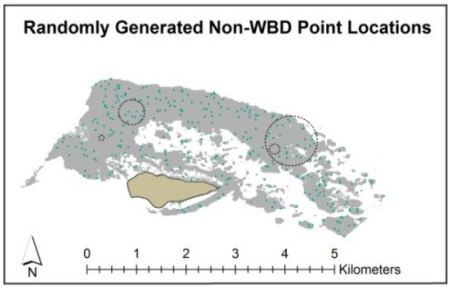
4) Next, I used *Hawth's Tools* to generate random point locations within this study area. These point locations were used to simulate the locations of transects containing *Acropora palmata* from Mayor's study. Since Mayor et al.'s study ultimately surveyed 375 transects, a total of 375 point locations were randomly generated for the Artificial dataset.

Creating the clustered WBD point locations

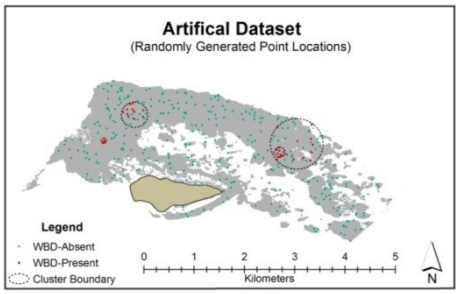


- Since Mayor's study found WBD in 44 of the 375 transects, 44 points were randomly generated for the artificial dataset.
- This was done by having *Hawth's Tools* generate 11 random points within each of the 4 cluster boundaries.
- Thus, Clusters 1-3 each had 11 WBD points within them, and Cluster 4 had 22 WBD points (since Cluster 2 was within its boundaries).
- *Note*: the 4 points used to define the center of these artificial clusters were not used to simulate a transect containing WBD.

Creating the artificial non-WBD point locations



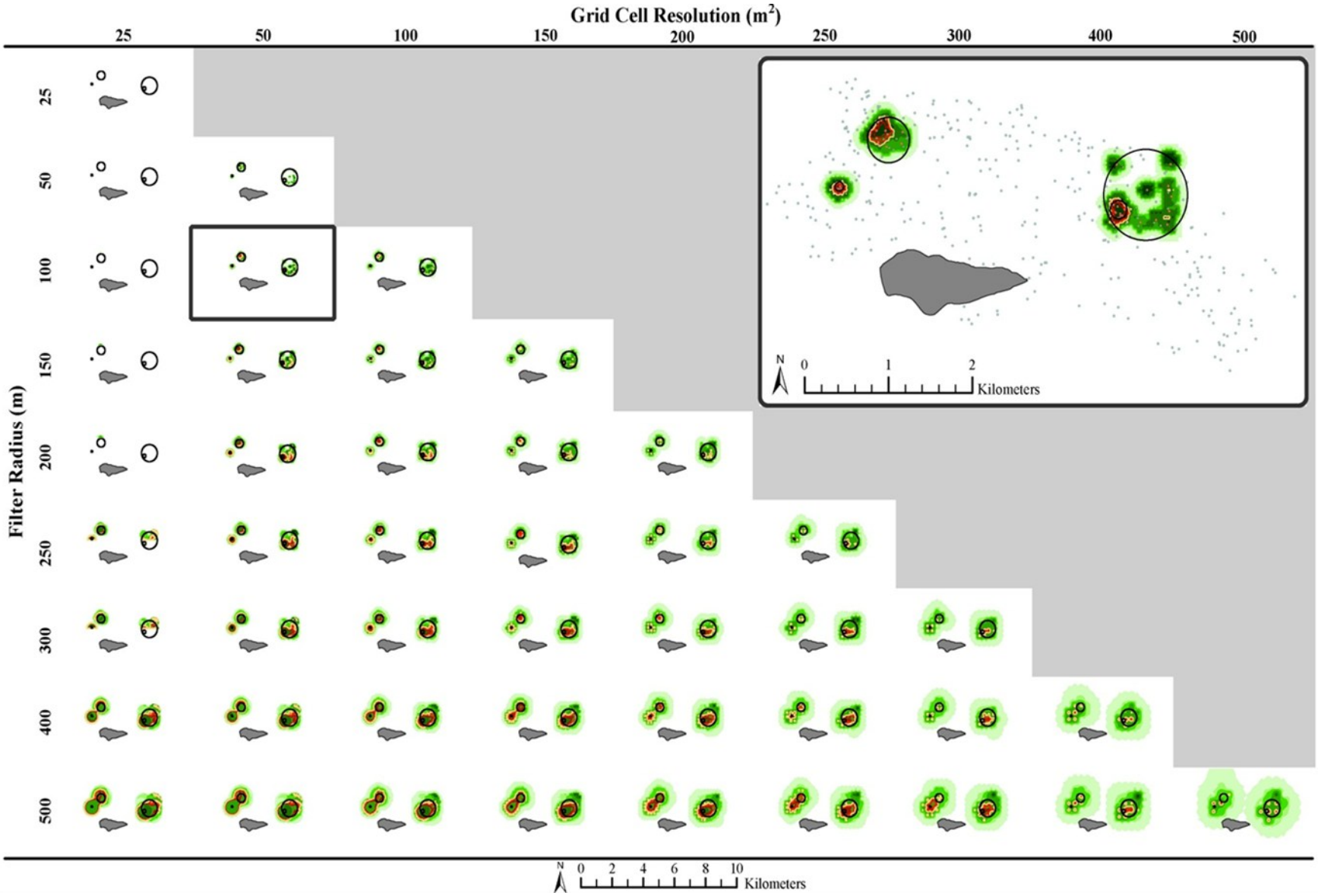
- *Hawth's Tools* was used to generate 331 random point locations within the study area.
- *Note*: Non-WBD points were allowed to fall within the WBD Cluster Boundaries, in order to simulate non-diseased, possibly resistant, corals during a WBD outbreak



- Thus this artificial dataset of randomly generated points can serve as a proxy for the transect locations in Mayor et al.'s 2006 study.

- The following types of Spatial Analysis will be performed on this artificial dataset in order to determine which types of analysis and parameter settings within these analyses will generate results that most closely proximate the 4 artificially defines WBD clusters.
- The analyses, and parameter settings within them, found to work best on the artifical dataset will then be applied to Mayor's actual data.
- In a sense this dataset will be used to calibrate the spatial analysis programs for Mayor's data. Thus, substantially increasing the probability that spatial patterns found in Mayor's data are real, and not just artifacts of the user defined parameter settings during the analysis [2, 5-7].

Calibrating the Spatial Parameters using the Artificial Cluster Dataset [2-3, 5-7]



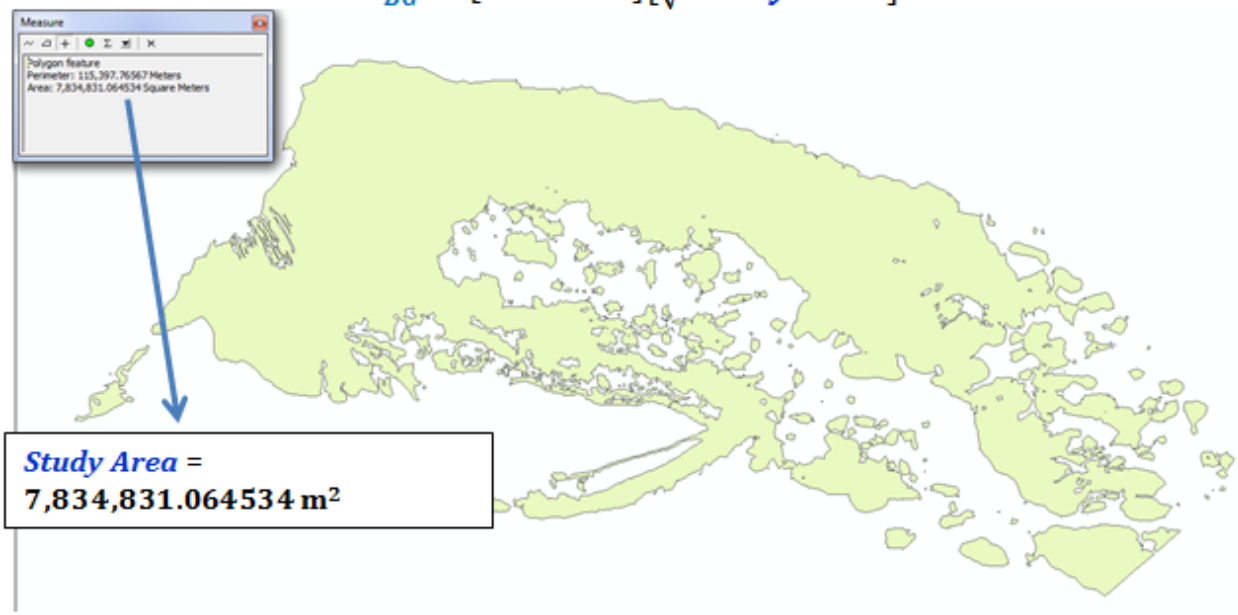
A.3 Plug-in Equation Bandwidth Estimation Methods

A.3.1 Default Search Radius used by ArcView's Kernel Density → h_{AV} [8-9]

$$h_{AV} = \frac{[\min(x, y)]}{30}$$

A.3.2 Bailey & Gatrell's h (h_{BG}) [8-10]

$$h_{BG} = [0.68n^{-0.2}][\sqrt{Study\ Area}]$$



$$h_{cases} = [0.68(\#\ of\ Case\ locations)^{-0.2}][\sqrt{Study\ Area}]$$
$$= [0.68(44)^{-0.2}]\left[\sqrt{7,834,831.064534m^2}\right]$$
$$= [0.319020][2799.0768m^2]$$
$$h_{cases} = \mathbf{892.96\ m^2}$$

$$h_{population} = [0.68(\#\ of\ Population\ locations)^{-0.2}][\sqrt{Study\ Area}]$$
$$= [0.68(375)^{-0.2}]\left[\sqrt{7,834,831.064534m^2}\right]$$
$$= [0.207827][2799.0768\,m^2]$$
$$h_{population} = \mathbf{581.72\ m^2}$$

A.3.3 Maximal Smoothing Bandwidth (h_{max}) [11-12]

$$h_{max} = [(1.144)(\sigma)]/\sqrt[5]{n}$$

Bandwidth estimated based on Numerator (Case) data

$$h_{max} = [(1.144)(\sigma)]/\sqrt[5]{44}$$

Artificially Clustered Points

$h_{max} = [(1.144)(\sigma)]/\sqrt[5]{n}$

$= [(1.144)(1720.54m)]/\sqrt[5]{44}$

$= [1968.30m]/\sqrt[5]{44}$

$h_{max} = 923.42m$

BUIS Transects with WBD present

$h_{max} = [(1.144)(\sigma)]/\sqrt[5]{n}$

$= [(1.144)(812.80m)]/\sqrt[5]{44}$

$= [1732.50m]/\sqrt[5]{44}$

$h_{max} = 812.80m$

Bandwidth estimated based on Denominator (Population) data

$$h_{max} = [(1.144)(\sigma)]/\sqrt[5]{375}$$

All Artificial Cluster Dataset Points

$h_{max} = [(1.144)(\sigma)]/\sqrt[5]{n}$

$= [(1.144)(1798.45m)]/\sqrt[5]{375}$

$= [2057.43m]/\sqrt[5]{375}$

$h_{max} = 628.81m$

All BUIS Transects

$h_{max} = [(1.144)(\sigma)]/\sqrt[5]{n}$

$= [(1.144)(1668.22m)]/\sqrt[5]{375}$

$= [1908.44m]/\sqrt[5]{375}$

$h_{max} = 583.27m$

A.3.4 Optimized Bandwidth (h_{opt}) [11-14]

$$h_{opt} = \left\{\left[\frac{2}{(3)(n)}\right]^{\frac{1}{4}}\right\}\{\sigma\}$$

Bandwidth estimated based on Numerator (Case) data

$$h_{opt} = \left\{\left[\frac{2}{(3)(44)}\right]^{\frac{1}{4}}\right\}\{\sigma\}$$

Artificially Clustered Points

$$h_{opt} = \left\{\left[\frac{2}{(3)(44)}\right]^{\frac{1}{4}}\right\}[\{1720.54m\}]$$

$$= \left\{\left[\frac{2}{132}\right]^{\frac{1}{4}}\right\}[\{1720.54m\}]$$

$$= \left\{[0.0\overline{15}]^{\frac{1}{4}}\right\}[\{1720.54m\}]$$

$$= \{0.35084397\}[\{1720.54m\}]$$

$$h_{opt} = 603.64m$$

BUIS Transects with WBD present

$$h_{opt} = \left\{\left[\frac{2}{(3)(44)}\right]^{\frac{1}{4}}\right\}\{1514.42m\}$$

$$= \left\{\left[\frac{2}{132}\right]^{\frac{1}{4}}\right\}[\{1514.42m\}]$$

$$= \left\{[0.0\overline{15}]^{\frac{1}{4}}\right\}[\{1514.42m\}]$$

$$= \{0.35084397\}[\{1514.42m\}]$$

$$h_{opt} = 531.33m$$

Bandwidth estimated based on Denominator (Population) data

$$h_{opt} = \left\{\left[\frac{2}{(3)(375)}\right]^{\frac{1}{4}}\right\}\{\sigma\}$$

All Artificial Cluster Dataset Points

$$h_{opt} = \left\{\left[\frac{2}{(3)(375)}\right]^{\frac{1}{4}}\right\}\{1798.45m\}$$

$$= \left\{\left[\frac{2}{1125}\right]^{\frac{1}{4}}\right\}[\{1798.45m\}]$$

$$= \left\{[0.00\overline{17}]^{\frac{1}{4}}\right\}[\{1798.45m\}]$$

$$= \{0.205338019\}[\{1798.45m\}]$$

$$h_{opt} = 369.29m$$

All BUIS Transects

$$h_{opt} = \left\{\left[\frac{2}{(3)(375)}\right]^{\frac{1}{4}}\right\}\{1668.22m\}$$

$$= \left\{\left[\frac{2}{1125}\right]^{\frac{1}{4}}\right\}[\{1668.22m\}]$$

$$= \left\{[0.00\overline{17}]^{\frac{1}{4}}\right\}[\{1668.22m\}]$$

$$= \{0.205338019\}[\{1668.22m\}]$$

$$h_{opt} = 342.55m$$

A.3.5 Reference Bandwidth (h_{ref}) [15-19]

h_{ref} is when h is the "optimum value with reference to a known standard distribution." [15-19]

"Since the HRE uses a **standard bivariate normal probability density function** to estimate the utilization distribution, h_{ref} is calculated as the square root of the mean variance in x (var_x) and y (var_y) co-ordinates divided by the sixth root of the number of points[20]"[15-16]

$$h_{ref} = n^{-1/6}\sqrt{(var_x + var_y)/2}$$

(Note: this is the non-standardized value of h_{ref})

Bandwidth estimated based on Numerator (Case) data

$$h_{ref} = 44^{-1/6}\sqrt{(var_x + var_y)/2}$$

Artificially Clustered Points

$$\begin{aligned} h_{ref} &= 44^{-1/6}\sqrt{(var_x + var_y)/2} \\ &= 44^{-1/6}\sqrt{(2774359.2500000 + 128438.5390625)/2} \\ &= 44^{-1/6}\sqrt{(2902797.7890625)/2} \\ &= 44^{-1/6}\sqrt{1451398.89453125} \\ &= 44^{-1/6}(1204.740177) \end{aligned}$$

$h_{ref} = 641.18686\ m$

BUIS Transects with WBD present

$$\begin{aligned} h_{ref} &= 44^{-1/6}\sqrt{(var_x + var_y)/2} \\ &= 44^{-1/6}\sqrt{(1796403.125 + 447127.25)/2} \\ &= 44^{-1/6}\sqrt{(2243530.375)/2} \\ &= 44^{-1/6}\sqrt{1121765.1875} \\ &= 44^{-1/6}(1059.134169) \end{aligned}$$

$h_{ref} = 563.69243\ m$

Bandwidth estimated based on Denominator (Population) data

$$h_{ref} = 375^{-1/6}\sqrt{(var_x + var_y)/2}$$

All Artificial Cluster Dataset Points

$$\begin{aligned} h_{ref} &= 375^{-1/6}\sqrt{(var_x + var_y)/2} \\ &= 375^{-1/6}\sqrt{(2670535.750000 + 561252.9375000)/2} \\ &= 375^{-1/6}\sqrt{(3231788.6875)/2} \\ &= 375^{-1/6}\sqrt{1615894.34375} \\ &= 375^{-1/6}(1271.178329) \end{aligned}$$

$h_{ref} = 473.37059\ m$

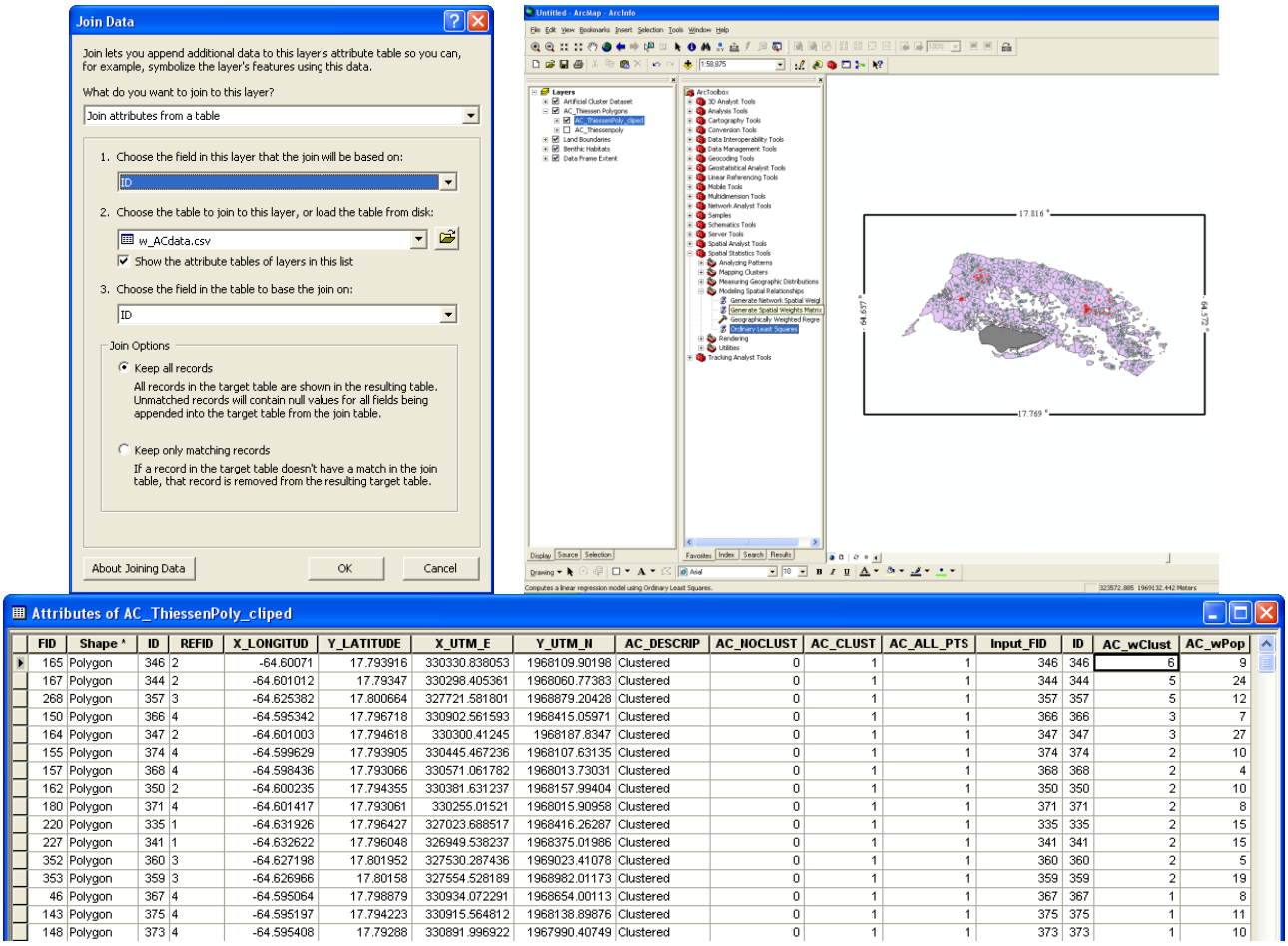
All BUIS Transects

$$\begin{aligned} h_{ref} &= 375^{-1/6}\sqrt{(var_x + var_y)/2} \\ &= 375^{-1/6}\sqrt{(2254249.0 + 525819.9375)/2} \\ &= 375^{-1/6}\sqrt{(2780068.9375)/2} \\ &= 375^{-1/6}\sqrt{1390034.46875} \\ &= 375^{-1/6}(1178.997230) \end{aligned}$$

$h_{ref} = 439.04352\ m$

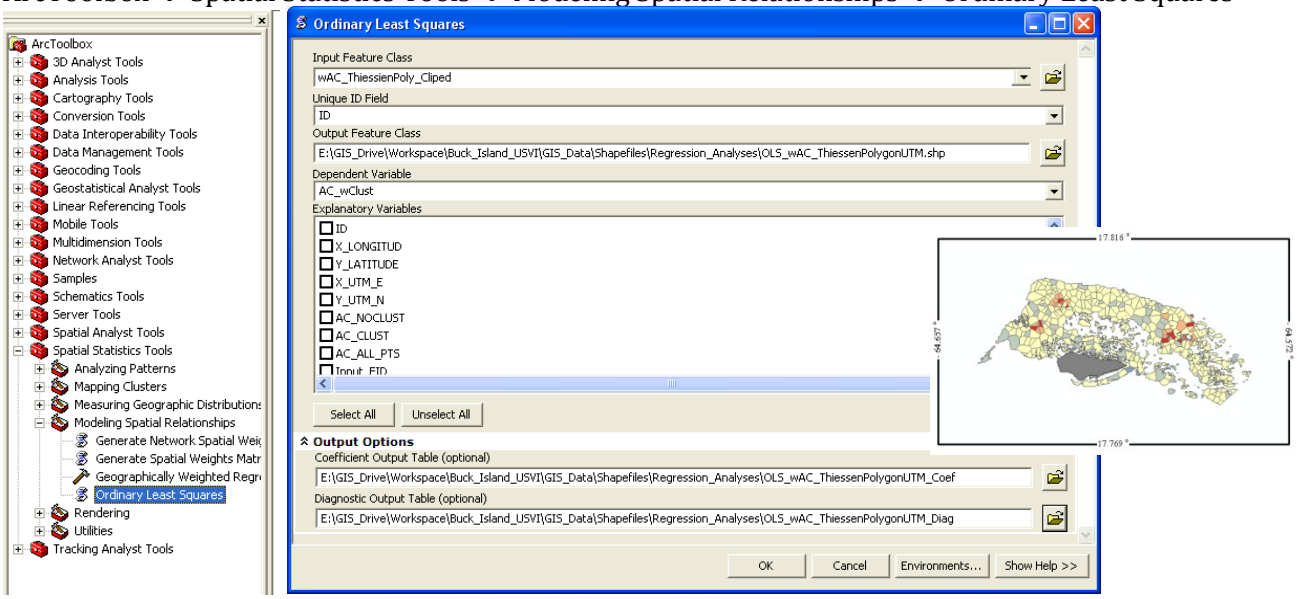
A.4 Regression-based Bandwidth Selection Criterion

A.4.1 Bandwidth Selection Criteria [21]

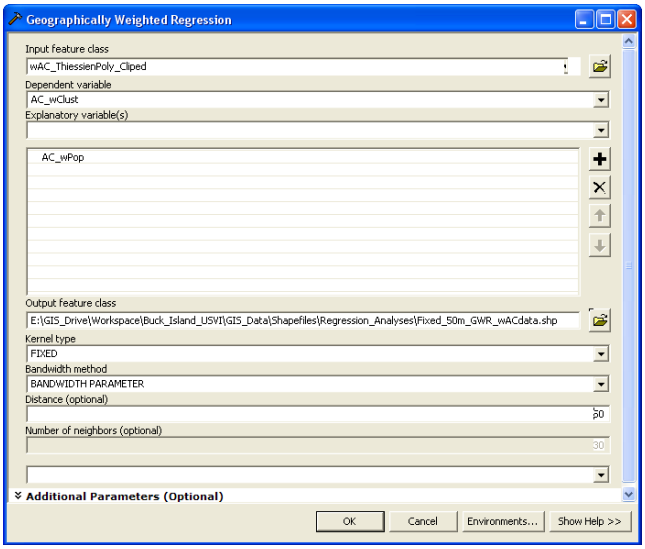


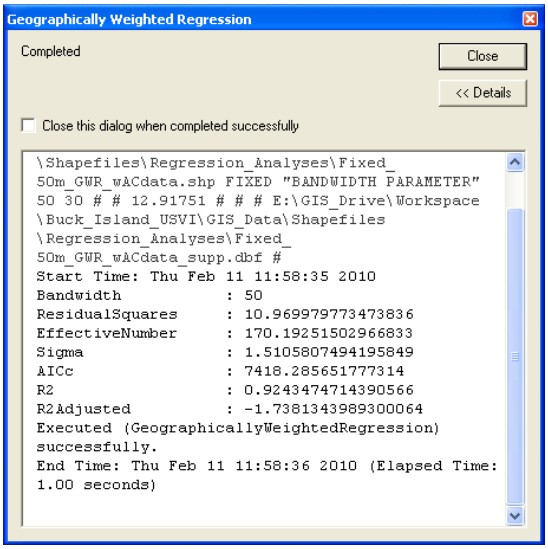
FID	Shape *	ID	REFID	X_LONGITUD	Y_LATITUDE	X_UTM_E	Y_UTM_N	AC_DESCRIP	AC_NOCLUST	AC_CLUST	AC_ALL_PTS	Input_FID	ID	AC_wClust	AC_wPop
165	Polygon	346	2	-64.60071	17.793916	330330.838053	1968109.90198	Clustered	0	1	1	346	346	6	9
167	Polygon	344	2	-64.601012	17.79347	330298.405361	1968060.77383	Clustered	0	1	1	344	344	5	24
268	Polygon	357	3	-64.625382	17.800664	327721.581801	1968879.20428	Clustered	0	1	1	357	357	5	12
150	Polygon	366	4	-64.595342	17.796718	330902.561593	1968415.05971	Clustered	0	1	1	366	366	3	7
164	Polygon	347	2	-64.601003	17.794618	330300.41245	1968187.8347	Clustered	0	1	1	347	347	3	27
155	Polygon	374	4	-64.599629	17.793905	330445.467236	1968107.63135	Clustered	0	1	1	374	374	2	10
157	Polygon	368	4	-64.598436	17.793066	330571.061782	1968013.73031	Clustered	0	1	1	368	368	2	4
162	Polygon	350	2	-64.600235	17.794355	330381.631237	1968157.99404	Clustered	0	1	1	350	350	2	10
180	Polygon	371	4	-64.601417	17.793061	330255.01521	1968015.90958	Clustered	0	1	1	371	371	2	8
220	Polygon	335	1	-64.631926	17.796427	327023.688517	1968416.26287	Clustered	0	1	1	335	335	2	15
227	Polygon	341	1	-64.632622	17.796048	326949.538237	1968375.01986	Clustered	0	1	1	341	341	2	15
352	Polygon	360	3	-64.627198	17.801952	327530.287436	1969023.41078	Clustered	0	1	1	360	360	2	5
353	Polygon	359	3	-64.626966	17.80158	327554.528189	1968982.01173	Clustered	0	1	1	359	359	2	19
46	Polygon	367	4	-64.595064	17.798879	330934.072291	1968654.00113	Clustered	0	1	1	367	367	1	8
143	Polygon	375	4	-64.595197	17.794223	330915.564812	1968138.89876	Clustered	0	1	1	375	375	1	11
148	Polygon	373	4	-64.595408	17.79288	330891.996922	1967990.40749	Clustered	0	1	1	373	373	1	10

ArcToolbox → Spatial Statistics Tools → Modeling Spatial Relationships → Ordinary Least Squares



→→→ Geographically Weighted Regression (GWR)





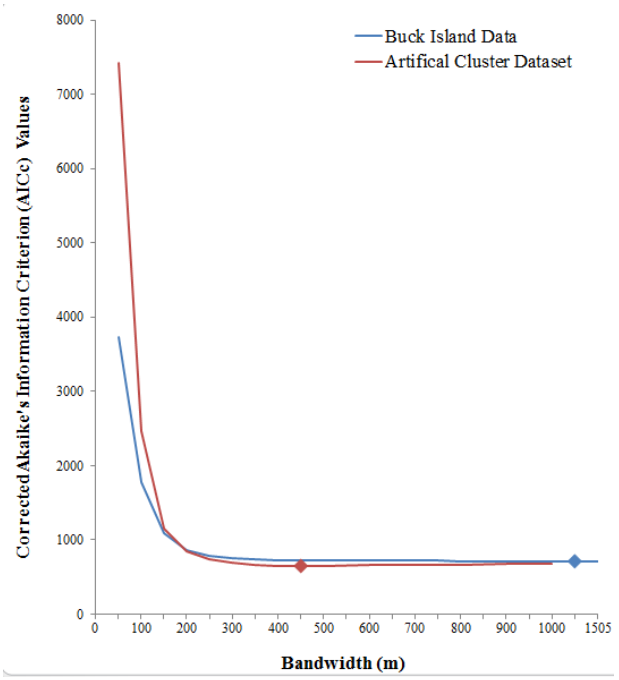
Repeat the above GWR analysis using distances of 100m-1,000 (in 50m increments)

KernelTYPE	bandMETHOD	Field_Type	BANDWIDTHm	AICc	GCV	LSC	ResidualSq	EffectiveN	Sigma	R2	R2Adjusted	Dependen_Y	Explanat_X
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	0										
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	50	3729.84201	636.3382174	51547.54634	1.57567	200.70873	0.41181	0.98652	0.69685	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	100	1776.12107	9441.963854	209970.8988	33.68875	283.68040	0.66879	0.76702	-0.10735	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	150	1091.83365	7942.509021	475437.0573	59.97372	225.17534	0.63269	0.61133	0.02978	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	200	863.71389	7309.319712	847946.022	77.59806	166.42548	0.60995	0.49711	0.09825	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	250	782.21258	7923.352783	1327497.793	91.51678	124.34404	0.60424	0.40691	0.11505	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	300	749.80474	9158.601802	1914092.37	102.39271	95.04870	0.60477	0.33642	0.11350	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	350	736.02979	10861.74842	2607729.752	110.73563	74.94763	0.60750	0.28235	0.10549	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	400	729.50024	12948.7748	3408409.942	117.14134	60.82096	0.61061	0.24084	0.09629	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	450	725.96452	15369.81246	4316132.937	122.07381	50.48950	0.61333	0.20887	0.08822	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	500	723.86051	18983.40831	5330898.738	125.90482	42.67889	0.61552	0.18405	0.08171	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	550	722.34677	21134.47909	6452707.345	128.86051	36.63078	0.61711	0.16489	0.07695	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	600	721.16170	24466.82126	7681558.759	131.15326	31.87565	0.61825	0.15003	0.07355	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	650	720.10477	28095.87491	9017452.978	132.92875	28.07444	0.61900	0.13853	0.07130	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	700	719.12340	32019.49412	10460390	134.31616	24.98848	0.61947	0.12954	0.06988	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	750	718.22293	36236.42736	12010369.84	135.41855	22.44966	0.61977	0.12239	0.06900	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	800	717.39973	40744.75446	13667392.47	136.30976	20.33150	0.61994	0.11662	0.06847	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	850	716.69784	45543.87075	15431457.92	137.05742	18.54547	0.62008	0.11177	0.06805	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	900	716.09112	50632.70663	17302566.17	137.69143	17.02284	0.62019	0.10766	0.06772	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	950	715.56891	56010.35178	19280717.22	138.23769	15.71170	0.62029	0.10412	0.06744	0	1
Fixed_Distance	BANDWIDTH PARAMETER	VALUE	1000	715.14423	61676.00793	21365911.09	138.72324	14.57226	0.62039	0.10097	0.06712	0	1
Fixed_Distance	AICc	VALUE	1298.85466	714.23301			140.93350	10.02992	0.62141	0.08665	0.06405	0	1
Fixed_Distance	CV	VALUE	1505.45035	714.40507			142.00576	8.19659	0.62221	0.07970	0.06165	0	1

A.4.2 The Corrected Akaike's Information Criterion (*AICc*) → h_{AICc} [21]

$$AIC_c = 2nlog_e(\hat{\sigma}) + nlog_e(2\pi) + n\left\{\frac{n+tr(S)}{n-2-tr(S)}\right\}$$

	AC_data	BUIS_data
bandwidth (m)	AIC_c	AIC_c
0		
50	7418.29	3729.84
100	2470.92	1776.12
150	1151.65	1091.83
200	852.52	863.71
250	739.23	782.21
300	687.79	749.80
350	664.00	736.03
400	654.21	729.50
450	651.45	725.96
500	652.19	723.86
550		722.35
600	657.49	721.16
650		720.10
700	664.06	719.12
750		718.22
800	670.36	717.40
850		716.70
900	676.17	716.09
950		715.57
1000	681.58	715.14
1299		714.23
1505		714.41



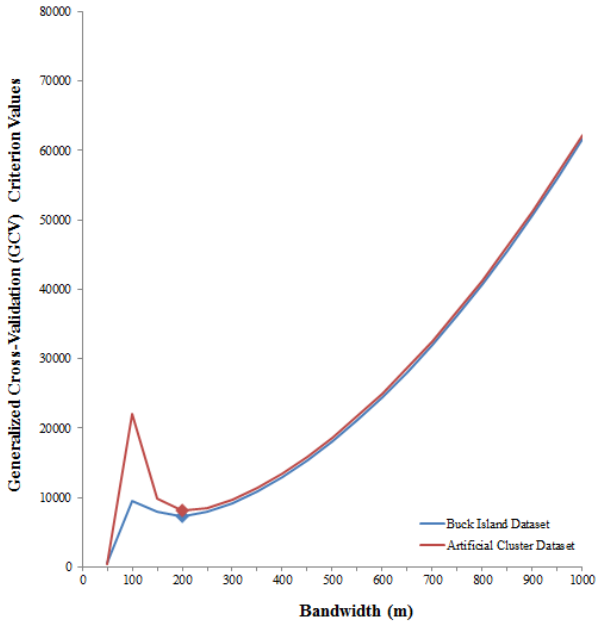
Total # of Locations	Dependent Field	Explanatory Field										
n	0	1	KernelTYPE	bandMETHOD	Field_Type	Bandwidthm	ResidualSq	EffectiveN	Sigma	AICc	R2	R2Adjusted
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	50	10.9699797735	170.1925150297	1.5105807494	7418.2856517773	0.9243474714	-1.7381343989
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	100	36.2445463104	315.1066277547	0.8438994807	2470.9192530568	0.7646447273	-0.6879344158
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	150	55.5306049171	240.8438319314	0.6482200850	1151.6524892501	0.6399640731	-0.0134477019
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	200	67.5887001824	177.3926357810	0.5848378671	852.5168642491	0.5619770052	0.1709792766
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	250	75.7936175943	133.3041279158	0.5599918437	739.2268614046	0.5088032871	0.2399225975
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	300	82.1222392445	103.3139543894	0.5497898592	687.7896449326	0.4677893039	0.2673646529
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	350	87.7260383509	82.1994760859	0.5473666510	663.9957218174	0.4314726880	0.2738086263
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	400	92.8944077327	66.9544853912	0.5491453877	654.2146674685	0.3979779673	0.2690812573
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	450	97.6182070935	55.6949057793	0.5529202424	651.4493989669	0.3673643775	0.2589979706
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	500	101.8903028714	47.1746526593	0.5575003721	652.1940734087	0.3396781492	0.2466709051
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	600	109.0266630189	35.3772074736	0.5665887481	657.4880317586	0.2934294444	0.2219091485
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	700	114.6522073864	27.7936907100	0.5746420544	664.0580797718	0.2569719036	0.1996328966
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	800	119.1574427142	22.6213239418	0.5815081117	670.3583968010	0.2277747647	0.1803924084
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	900	122.8828151192	18.9175353237	0.5874491535	676.1723747833	0.2036316938	0.1635596356
375	AC_wClust	AC_wPop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	1000	126.0717632591	16.1609378508	0.5927329318	681.5766505334	0.1829650349	0.1484453361
375	AC_wClust	AC_wPop	Fixed_Distance	AICc	VALUE	459.54186	98.4743670886	53.8869474503	0.5537740740	651.4125765708	0.3618158500	0.2567076604
375	AC_wClust	AC_wPop	Fixed_Distance	CV	VALUE	440.82151	96.7876840486	57.5257906327	0.5521484075	651.6308907147	0.3727467593	0.2610652926

Total # of Locations (Transects, Sites, etc)		Dependent Field	Explanatory Field									
n	description of n	0	1	KernelTYPE	bandMETHOD	Field_Type	Bandwidthm	ResidualSq	EffectiveN	Sigma	AICc	R2	R2Adjusted
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	50	1.57567	200.70873	0.41181	3729.84201	0.98652	0.69685
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	100	33.68875	283.68040	0.66879	1776.12107	0.76702	-0.10735
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	150	59.97372	225.17534	0.63269	1091.83365	0.61133	0.02978
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	200	77.59806	166.42548	0.60995	863.71389	0.49711	0.09825
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	250	91.51678	124.34404	0.60424	782.21258	0.40691	0.11505
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	300	102.39271	95.04870	0.60477	749.80474	0.33642	0.11350
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	350	110.73563	74.94763	0.60750	736.02979	0.28235	0.10549
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	400	117.14134	60.82096	0.61061	729.50024	0.24084	0.09629
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	450	122.07381	50.48950	0.61333	725.96452	0.20887	0.08822
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	500	125.90482	42.67889	0.61552	723.86051	0.18405	0.08171
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	550	128.86051	36.63078	0.61711	722.34677	0.16489	0.07695
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	600	131.15326	31.87565	0.61825	721.16170	0.15003	0.07355
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	650	132.92875	28.07444	0.61900	720.10477	0.13853	0.07130
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	700	134.31616	24.98848	0.61947	719.12340	0.12954	0.06988
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	750	135.41855	22.44966	0.61977	718.22293	0.12239	0.06900
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	800	136.30976	20.33150	0.61994	717.39973	0.11662	0.06847
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	850	137.05742	18.54547	0.62008	716.69784	0.11177	0.06805
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	900	137.69143	17.02284	0.62019	716.09112	0.10766	0.06772
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	950	138.23769	15.71170	0.62029	715.56891	0.10412	0.06744
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	BANDWIDTH PARAMETER	VALUE	1000	138.72324	14.57226	0.62039	715.14423	0.10097	0.06712
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	AICc	VALUE	1298.85466	140.93350	10.02992	0.62141	714.23301	0.08665	0.06405
375	25 by 10m transects	w_WBD	W_Pop	Fixed_Distance	CV	VALUE	1505.45035	142.00576	8.19659	0.62221	714.40507	0.07970	0.06165

A.4.3 The Generalized Cross-Validation Criterion (*GCV*) → h_{GCV} [21]

$$GCV = n\sum_{i=1}^{n}[y_i - \hat{y}_i(b)]^2/(n-v_1)^2$$

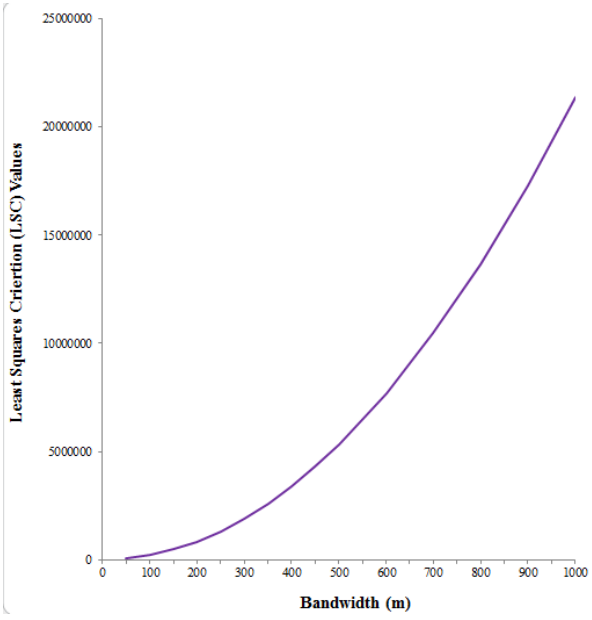
bandwidth (m)	AC_data GCV	BUIS_data GCV
0		
50	460.84	636.34
100	21949.91	9441.96
150	9906.11	7942.51
200	8143.16	7309.32
250	8521.71	7923.35
300	9724.33	9158.60
350	11406.44	10861.75
400	13469.56	12948.77
450	15875.02	15369.81
500	18601.44	18101.55
550		21134.48
600	24973.93	24466.82
650		28095.87
700	32538.98	32019.49
750		36236.43
800	41276.01	40744.75
850		45543.87
900	51172.97	50632.71
950		56010.35
1000	62223.33	61676.01
1299		
1505		



A.4.4 The Least Squares Criterion (*LSC*) → h_{LSC} [21]

$$LSC = \sum_{i=1}^{n}[y_i - \hat{y}_i(b)]^2$$

bandwidth (m)	AC_data LSC	BUIS_data LSC
0		
50	51547.55	51547.55
100	209970.90	209970.90
150	475437.06	475437.06
200	847946.02	847946.02
250	1327497.79	1327497.79
300	1914092.37	1914092.37
350	2607729.75	2607729.75
400	3408409.94	3408409.94
450	4316132.94	4316132.94
500	5330898.74	5330898.74
550		6452707.35
600	7681558.76	7681558.76
650		9017452.98
700	10460390.00	10460390.00
750		12010369.84
800	13667392.47	13667392.47
850		15431457.92
900	17302566.17	17302566.17
950		19280717.22
1000	21365911.09	21365911.09
1299		
1505		



A.4.5 The Least Squares Cross-Validation (*LSCV*) Criterion → h_{CV} [15-17, 20-23]

The **Least-Squares cross-validation (*LSCV*)** calculates the smoothing parameter (h_{cv}) by finding the h that minimizes the **mean integrated square error (*MISE*)** by minimizing the score function ***CV(h)*** for the estimated error between the true density function & kernel density estimate [15-17, 20].

$$CV(h) = \frac{1}{\pi n h^2} + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\left(\frac{1}{4\pi} e^{-D_{ij}/4} - \frac{1}{\pi} e^{-D_{ij}/2}\right)}{n^2 h^2}$$

"Where the distance between pairs of points (D_{ij}) is calculated as [16]:

$$D_{ij} = \left[\frac{(X_i - X_j)}{h}\right]^2 = \frac{\left(X_{i(x)} - X_{j(x)}\right)^2 + \left(X_{i(y)} - X_{j(y)}\right)^2}{h^2}$$

The minimum value of ***CV(h)*** is found by testing values of h between $0.01h_{ref}$ and $h_{ref,}$ selected using a 'Golden Section Search' algorithm [22]. The resulting smoothing parameter that minimizes the score function is called h_{cv}." [15-17]

Note: "In situations where the utilization distribution is not unimodal, the *LSCV* method has been shown to overcome the problem of over-smoothing associated with the use of h_{ref} [20]. However, the *LSCV* method is not always successful in finding a smoothing parameter that will minimize the *MISE*. In these cases, the HRE will report a low value for h_{cv} that is near the smallest value that can be tested (i.e., 0.01 h_{ref}). This will result in a utilization distribution that is seriously under-smoothed. Indeed, the *LSCV* method has a propensity to show structure in the data when none exists [23]."

A.4.6 The Biased Cross-Validation (*BCV*) Criterion → h_{BCV2} [15-17, 20-23]

The **biased cross-validation** (***BCV***) technique that may strike a balance between the tendency of h_{ref} to oversmooth and h_{cv} to undersmooth. "In contrast with the *LSCV* method, *BCV* attempts to find a value for h that minimizes an estimate of the **asymptotic mean integrated square error** (***AMISE***). *AMISE* is a large sample (e.g., n>50) approximation of the *MISE* [25]. Thus, it also provides an **estimate of the difference between** the **true density function** and the **kernel density estimate**. However, it is computationally faster and easier to calculate than *MISE* and provides a more direct indication of the performance of h values [23-25]. In the HRE, the function to be minimized is [16-17, 23]:

$$BCV2(h) = \frac{1}{4\pi n h^2 (n-1)} + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{(D_{ij}^2 - 8D_{ij} + 8)e^{-D_{ij}^2/2}}{8(n-1)(n-2)h^2\pi}$$

where the distance between pairs of points (D_{ij}) is again calculated as:

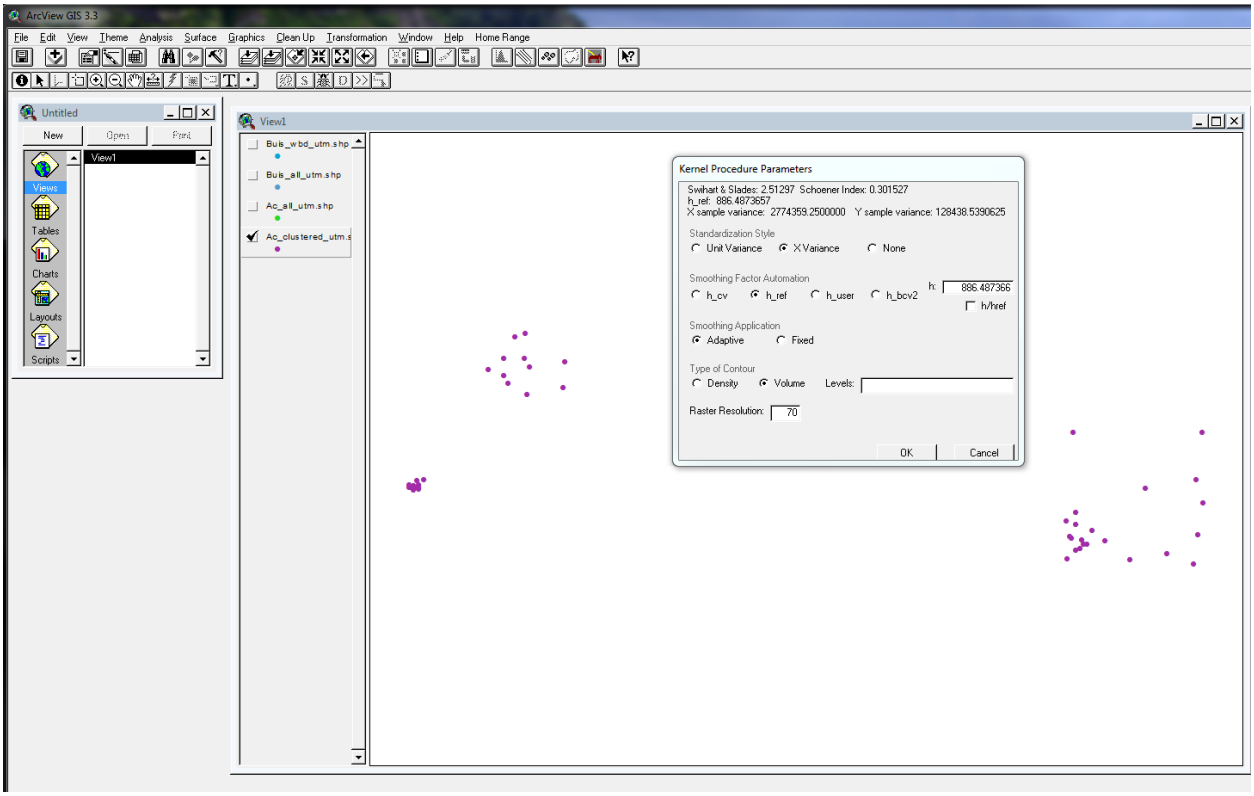
$$D_{ij} = \left[\frac{(X_i - X_j)}{h}\right]^2 = \frac{\left(X_{i(x)} - X_{j(x)}\right)^2 + \left(X_{i(y)} - X_{j(y)}\right)^2}{h^2}$$

Similar to the *LSCV* method, values of h between $0.01h_{ref}$ and $h_{ref,}$ are selected for testing using a "**Golden Section Search**" algorithm [22]. The resulting smoothing parameter is called h_{bcv2}.

Simulation studies show the *BCV* method performs quite well and with reasonable variability in comparisons with the *LSCV* and reference methods [23]. However, the *BCV* method has not been investigated in the context of home range estimation." [15-17].

A.4.7 h_{CV}, h_{ref}, & h_{BCV2} calculated by the HRE extension in ArcView 3.3x [16]

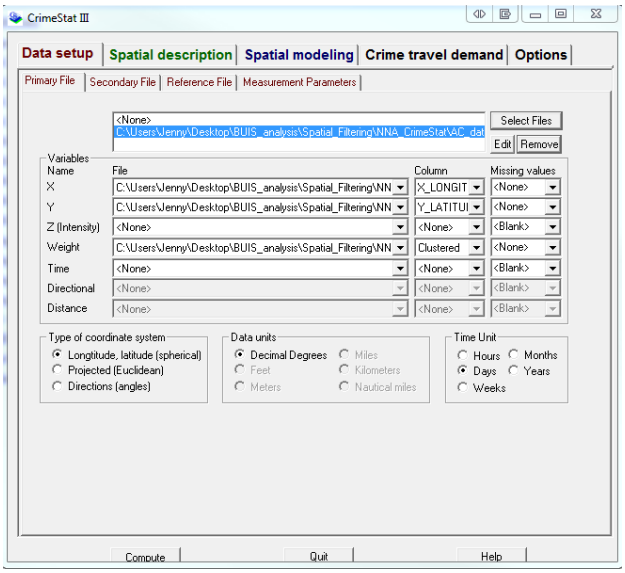
			Case Data		Population Data	
			AC Clustered Pts	WBD Transects	AC all points	All BUIS Transects
		Swihart & Slades	2.51297	0.05032652	0.413063	0.083844
		Schoener Index	0.301527	2.10898	1.71616	2.03465
		X sample variance	2774359.2500000	1796403.125	2670535.7500	2254249.000
		Y sample variance	128438.5390625	447127.25	561252.93750	525819.9375
Standardization Style	Unit Variance	h_{cv}	0.054559	0.134373	0.099496	0.141936
		h_{ref}	0.5322201	0.532220	0.3723872	0.372387
		h_{bcv2}	0.054559	0.134373	0.099496	0.141936
	X Variance	h_{cv}	90.876038	180.099686	162.594131	213.105682
		h_{ref}	886.4873657	713.334351	608.546814	559.108032
		h_{bcv2}	90.876038	180.099686	162.594131	213.105682
	None	h_{cv}	65.729668	138.610870	100.097137	132.427002
		h_{ref}	641.186829	563.692444	473.370575	439.043518
		h_{bcv2}	65.729668	138.610870	100.097137	132.427002
Standardization Style= Unit or X Variance		$h = h_{cv} / h_{ref}$	0.102513	0.252476	0.267184	0.3871153
		$h = h_{ref}/h_{ref}$	1.000000	1.000000	1.000000	1.0000000
		$h = h_{bcv2}/h_{ref}$	0.102513	0.252476	0.267184	0.381153
Standardization Style = None		$h = h_{cv} / h_{ref}$	0.102513	0.245898	0.211456	0.301626
		$h = h_{ref}/h_{ref}$	1.000000	1.0000000	1.000000	1.000000
		$h = h_{bcv2}/h_{ref}$	0.102513	0.245898	0.211456	0.301626

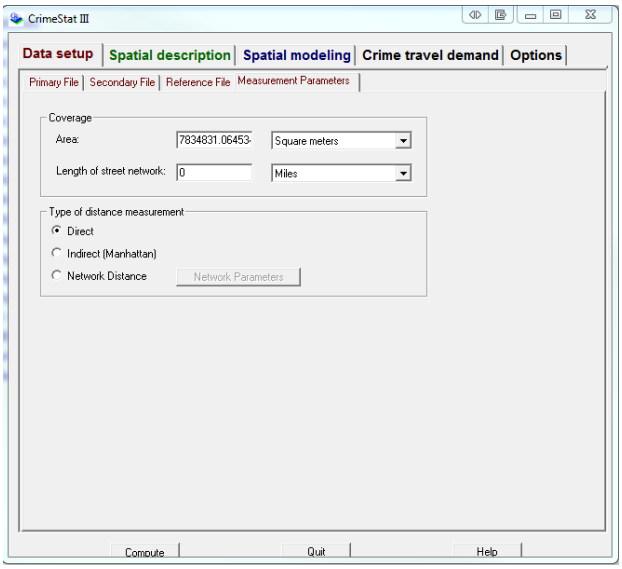


A.5. Nearest Neighbor Analysis Bandwidth Selection Methods

A.5.1 Mean Nearest Neighbor Analysis (*Nna*) → h_{Nna} [1, 8-9, 27]

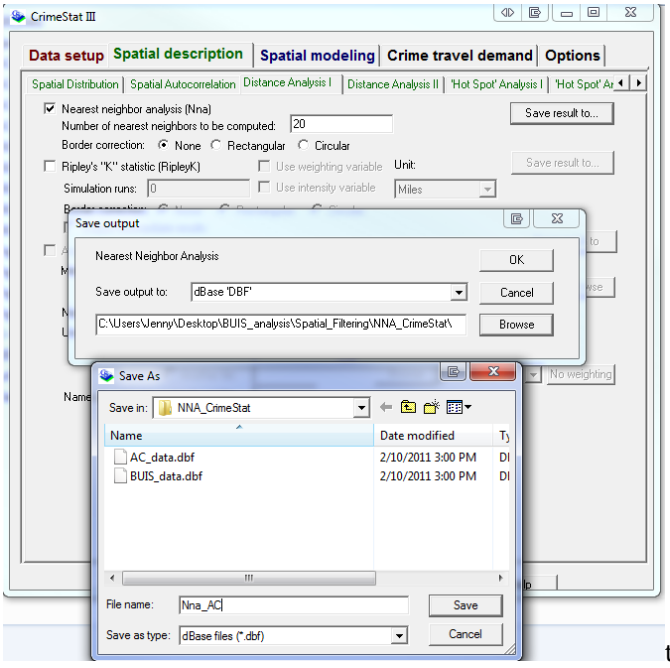
$$Nna = \sum_{i=1}^{n} \left(\frac{min(d_{ij})}{n} \right)$$





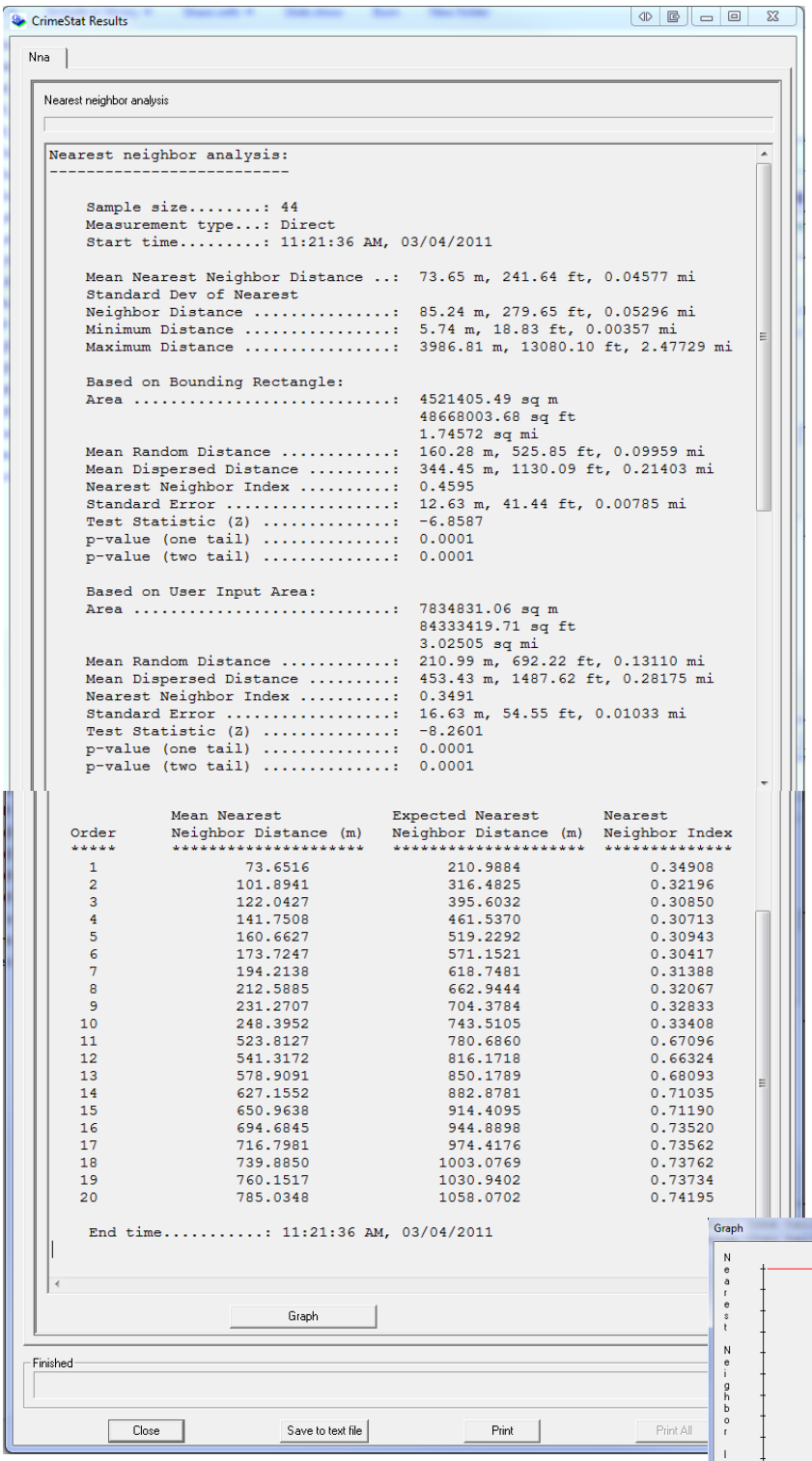
Click on the **Spatial description** tab & then click on the **Distance Analysis I** sub-tab

Check the **Nearest neighbor analysis (Nna)** box & set the # of nearest neighbors to be computed to "20"
("this will generate the mean nearest neighbor distance values for 1 to 20K orders" [1, p. 158]



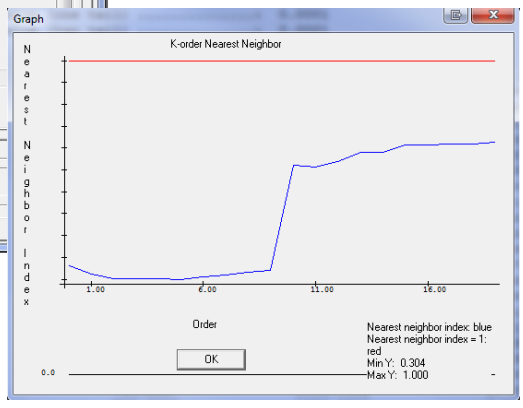
then hit "Compute"

The Results for the Nna of the Artificially Clustered point data

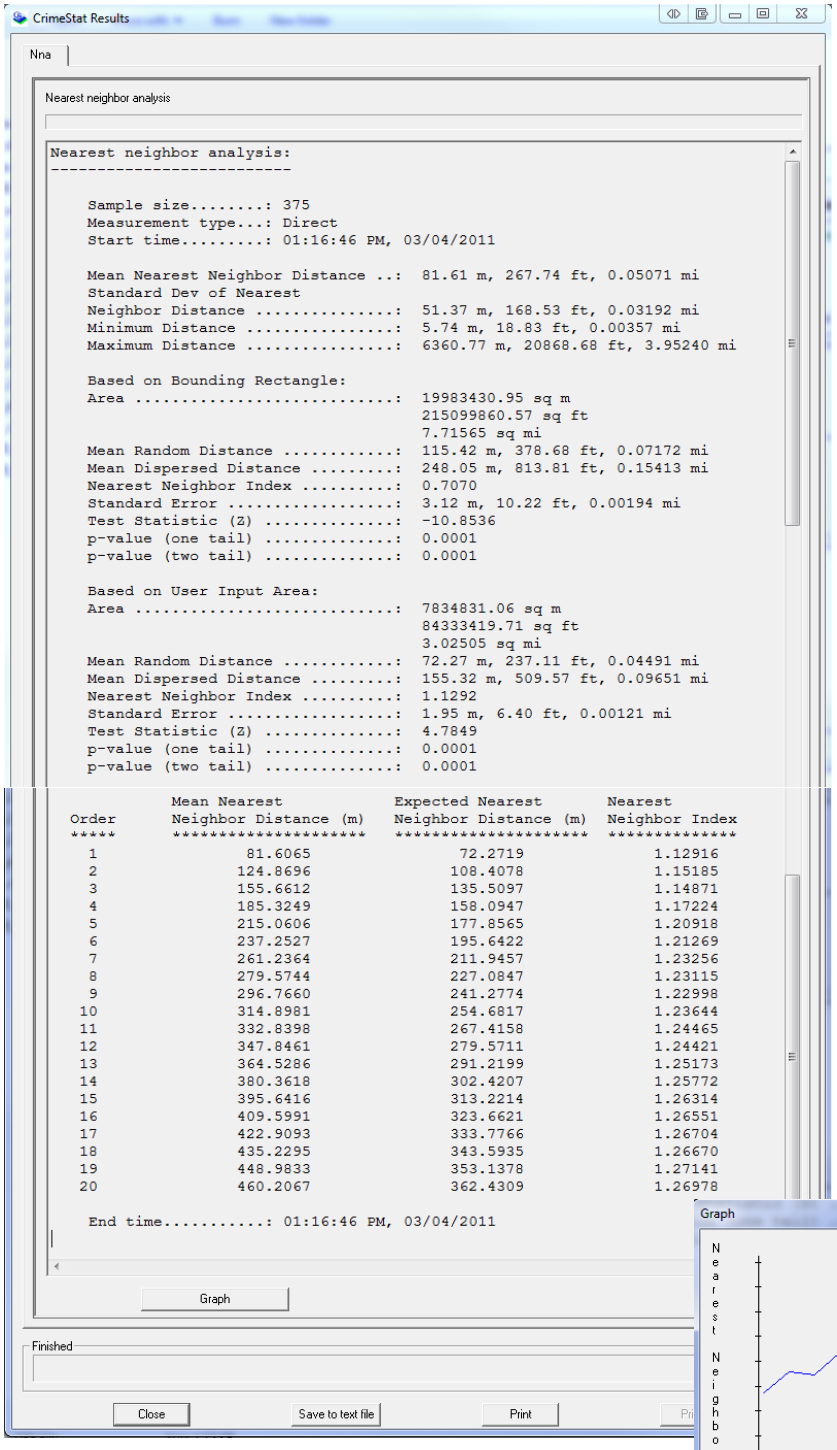


Results file saved as:
"Nna_AC_clustered.dbf"

```
Nearest neighbor analysis:
--------------------------

    Sample size.....: 44
    Measurement type...: Direct
    Start time.....: 11:21:36 AM, 03/04/2011

    Mean Nearest Neighbor Distance ..:  73.65 m, 241.64 ft, 0.04577 mi
    Standard Dev of Nearest
    Neighbor Distance ..............:  85.24 m, 279.65 ft, 0.05296 mi
    Minimum Distance ...............:  5.74 m, 18.83 ft, 0.00357 mi
    Maximum Distance ...............:  3986.81 m, 13080.10 ft, 2.47729 mi

    Based on Bounding Rectangle:
    Area ...........................:  4521405.49 sq m
                                       48668003.68 sq ft
                                       1.74572 sq mi
    Mean Random Distance ...........:  160.28 m, 525.85 ft, 0.09959 mi
    Mean Dispersed Distance ........:  344.45 m, 1130.09 ft, 0.21403 mi
    Nearest Neighbor Index .........:  0.4595
    Standard Error .................:  12.63 m, 41.44 ft, 0.00785 mi
    Test Statistic (Z) .............:  -6.8587
    p-value (one tail) .............:  0.0001
    p-value (two tail) .............:  0.0001

    Based on User Input Area:
    Area ...........................:  7834831.06 sq m
                                       84333419.71 sq ft
                                       3.02505 sq mi
    Mean Random Distance ...........:  210.99 m, 692.22 ft, 0.13110 mi
    Mean Dispersed Distance ........:  453.43 m, 1487.62 ft, 0.28175 mi
    Nearest Neighbor Index .........:  0.3491
    Standard Error .................:  16.63 m, 54.55 ft, 0.01033 mi
    Test Statistic (Z) .............:  -8.2601
    p-value (one tail) .............:  0.0001
    p-value (two tail) .............:  0.0001

              Mean Nearest           Expected Nearest       Nearest
  Order       Neighbor Distance (m)  Neighbor Distance (m)  Neighbor Index
  *****       ********************   ********************   **************
    1               73.6516                210.9884              0.34908
    2              101.8941                316.4825              0.32196
    3              122.0427                395.6032              0.30850
    4              141.7508                461.5370              0.30713
    5              160.6627                519.2292              0.30943
    6              173.7247                571.1521              0.30417
    7              194.2138                618.7481              0.31388
    8              212.5885                662.9444              0.32067
    9              231.2707                704.3784              0.32833
   10              248.3952                743.5105              0.33408
   11              523.8127                780.6860              0.67096
   12              541.3172                816.1718              0.66324
   13              578.9091                850.1789              0.68093
   14              627.1552                882.8781              0.71035
   15              650.9638                914.4095              0.71190
   16              694.6845                944.8898              0.73520
   17              716.7981                974.4176              0.73562
   18              739.8850               1003.0769              0.73762
   19              760.1517               1030.9402              0.73734
   20              785.0348               1058.0702              0.74195

    End time.....: 11:21:36 AM, 03/04/2011
```



The Results for the Nna of <u>all 375 points in the Artificially Clustered dataset</u>



Results file saved as:
"Nna_AC_all.dbf"

The Results for the Nna of the WBD transect data

Results file saved as:
"Nna_BUIS_WBD.dbf"

CrimeStat Results

Nna

Nearest neighbor analysis

```
Nearest neighbor analysis:
--------------------------

    Sample size.........: 44
    Measurement type...: Direct
    Start time.........: 01:07:31 PM, 03/04/2011

    Mean Nearest Neighbor Distance ..:  183.64 m, 602.49 ft, 0.11411 mi
    Standard Dev of Nearest
    Neighbor Distance ..............:  159.87 m, 524.52 ft, 0.09934 mi
    Minimum Distance ...............:  42.33 m, 138.87 ft, 0.02630 mi
    Maximum Distance ...............:  4924.08 m, 16155.12 ft, 3.05968 mi

    Based on Bounding Rectangle:
    Area ...........................:  12809689.74 sq m
                                       137882352.83 sq ft
                                       4.94585 sq mi
    Mean Random Distance ...........:  269.78 m, 885.11 ft, 0.16763 mi
    Mean Dispersed Distance ........:  579.78 m, 1902.16 ft, 0.36026 mi
    Nearest Neighbor Index .........:  0.6807
    Standard Error .................:  21.26 m, 69.75 ft, 0.01321 mi
    Test Statistic (Z) .............:  -4.0519
    p-value (one tail) .............:  0.0001
    p-value (two tail) .............:  0.0001

    Based on User Input Area:
    Area ...........................:  7834831.06 sq m
                                       84333419.71 sq ft
                                       3.02505 sq mi
    Mean Random Distance ...........:  210.99 m, 692.22 ft, 0.13110 mi
    Mean Dispersed Distance ........:  453.43 m, 1487.62 ft, 0.28175 mi
    Nearest Neighbor Index .........:  0.8704
    Standard Error .................:  16.63 m, 54.55 ft, 0.01033 mi
    Test Statistic (Z) .............:  -1.6449
    p-value (one tail) .............:  0.1
    p-value (two tail) .............:          n.s.
```

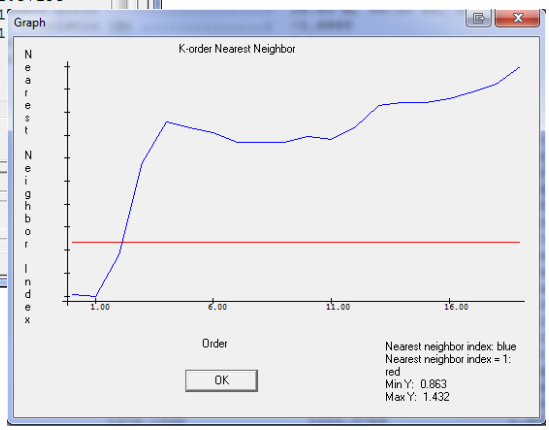
Order	Mean Nearest Neighbor Distance (m)	Expected Nearest Neighbor Distance (m)	Nearest Neighbor Index
1	183.6398	210.9884	0.87038
2	273.2154	316.4825	0.86329
3	382.8620	395.6032	0.96779
4	551.7344	461.5370	1.19543
5	673.5442	519.2292	1.29720
6	732.7862	571.1521	1.28300
7	786.1926	618.7481	1.27062
8	826.5202	662.9444	1.24674
9	877.9950	704.3784	1.24648
10	927.1383	743.5105	1.24697
11	983.3955	780.6860	1.25966
12	1023.2696	816.1718	1.25374
13	1090.8860	850.1789	1.28313
14	1180.9326	882.8781	1.33759
15	1229.0948	914.4095	1.34414
16	1269.9080	944.8898	1.34397
17	1318.9956	974.4176	1.35362
18	1376.1532	1003.0769	1.37193
19	1434.1930	1030.9402	1
20	1515.2433	1058.0702	1

```
    End time...........: 01:07:31 PM, 03/04/2011
```

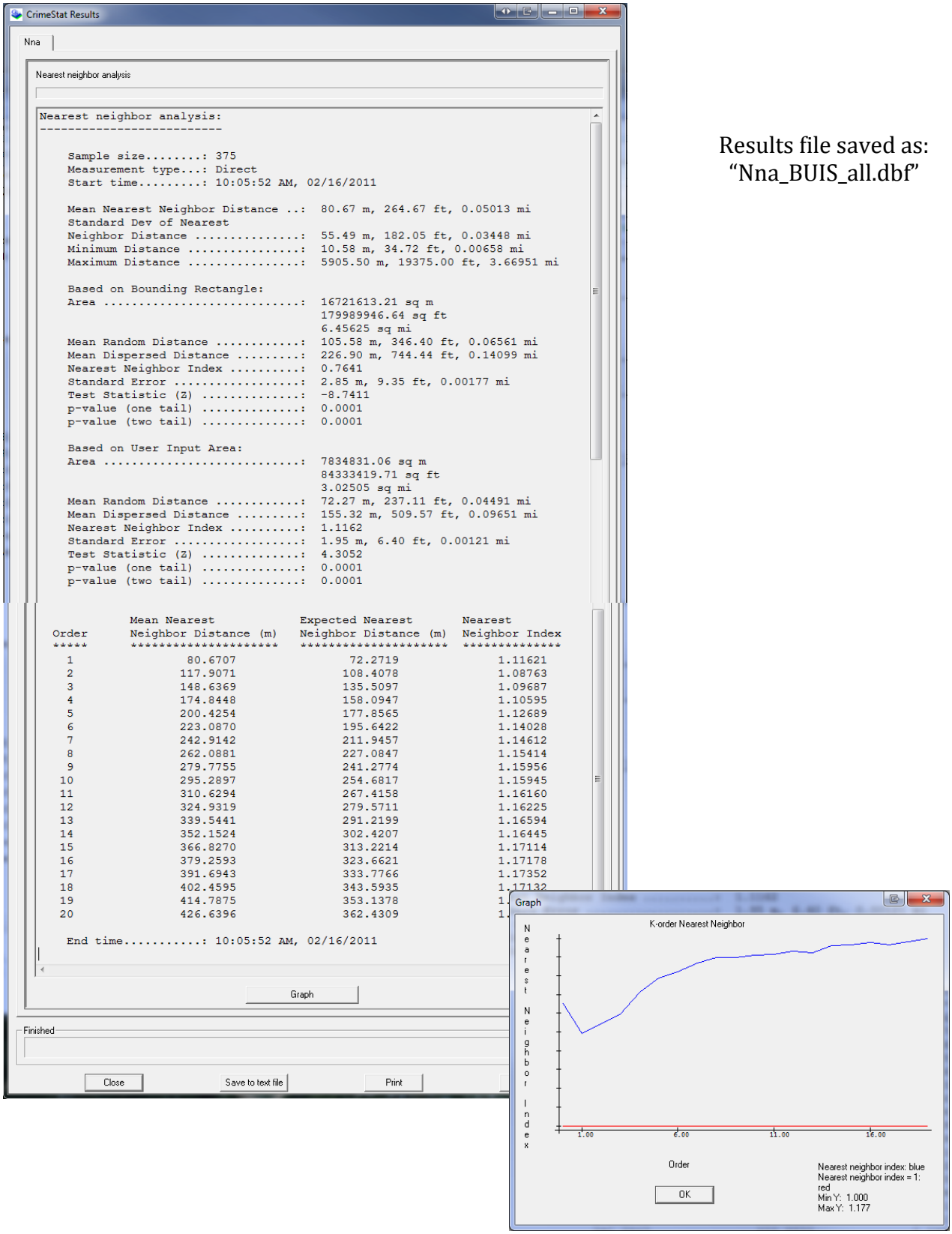
Graph

Finished

Close | Save to text file | Print



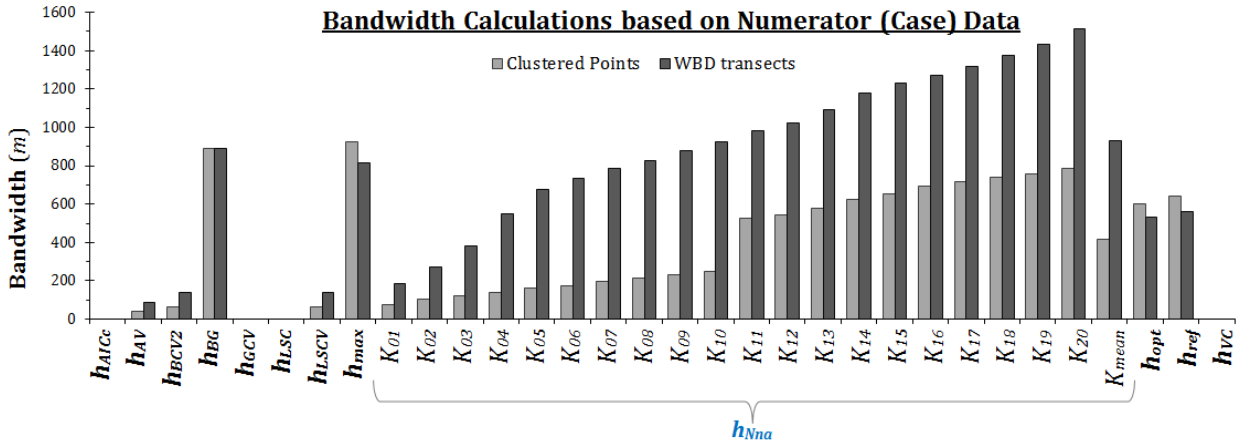
The Results for the Nna of <u>all 375 BUIS transects</u>

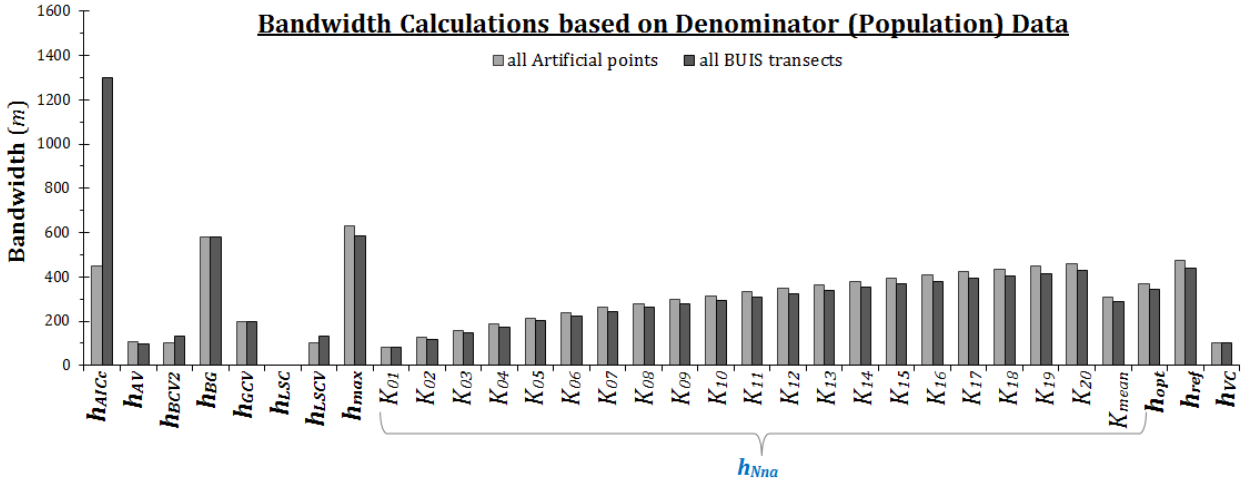


Results file saved as:
"Nna_BUIS_all.dbf"

A.6 Summary of Estimated Bandwidths

	Bandwidth Calculation based on			
	Numerator data ($n = 44$)		**Denominator data** ($n = 375$)	
	Clustered Points	**WBD transects**	**all Artificial points**	**All BUIS transects**
h_{AICc}			450.00	1298.85
h_{AV}	38.67	88.36	106.83	95.79
h_{BCV2}	65.73	138.61	100.10	132.43
h_{BG}	892.96	892.96	581.72	581.72
h_{GCV}			200.00	200.00
h_{LSC}				
h_{LSCV}	65.73	138.61	100.10	132.43
h_{max}	923.42	812.80	628.81	583.27
h_{Nna}, K_1	73.65	183.64	81.61	80.67
h_{Nna}, K_2	101.89	273.22	124.87	117.91
h_{Nna}, K_3	122.04	382.86	155.66	148.64
h_{Nna}, K_4	141.75	551.73	185.32	174.84
h_{Nna}, K_5	160.66	673.54	215.06	200.43
h_{Nna}, K_6	173.72	732.79	237.25	223.09
h_{Nna}, K_7	194.21	786.19	261.24	242.91
h_{Nna}, K_8	212.59	826.52	279.57	262.09
h_{Nna}, K_9	231.27	877.99	296.77	279.78
h_{Nna}, K_{10}	248.40	927.14	314.90	295.29
h_{Nna}, K_{11}	523.81	983.40	332.84	310.63
h_{Nna}, K_{12}	541.32	1023.27	347.85	324.93
h_{Nna}, K_{13}	578.91	1090.89	364.53	339.54
h_{Nna}, K_{14}	627.16	1180.93	380.36	352.15
h_{Nna}, K_{15}	650.96	1229.09	395.64	366.83
h_{Nna}, K_{16}	694.68	1269.91	409.60	379.26
h_{Nna}, K_{17}	716.80	1319.00	422.91	391.69
h_{Nna}, K_{18}	739.88	1376.15	435.23	402.46
h_{Nna}, K_{19}	760.15	1434.19	448.98	414.79
h_{Nna}, K_{20}	785.03	1515.24	460.21	426.64
h_{Nna}, K_{mean}	413.95	931.88	307.52	286.73
h_{opt}	603.64	531.33	369.29	342.55
h_{ref}	641.19	563.69	473.37	439.04
h_{VC}			100.00	100.00





A.7 Literature Cited in Appendix A

1. Chainey S, Ratcliffe J (2005) GIS and Crime Mapping. John Wiley & Sons, Ltd, Chichester, West Sussex, England
2. Lentz JA, Curtis AJ, Sammarco PW, Mayor PA (2008) Applying Medical Geography to Identify Spatial Hotspots of Coral Diseases. 11th International Coral Reef Symposium (ICRS), Fort Lauderdale, FL
3. Mayor PA, Rogers CS, Hillis-Starr ZM (2006) Distribution and abundance of elkhorn coral, *Acropora palmata*, and prevalence of White-Band disease at Buck Island Reef National Monument, St. Croix, US Virgin Islands. Coral Reefs 25:239-242
4. Wiegand T, Gunatilleke S, Gunatilleke N, Okuda T (2007) Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. Ecology 88:3088-3102
5. Cai Q, Rushton G, Bhaduri B (2011) Validation tests of an improved kernel density estimation method for identifying disease clusters. Journal of Geographical Systems:1-22
6. Lentz JA, Curtis AJ, Mayor PA (2009) An Examination of how GIS and spatial analysis can be used to better understand coral health Graduate Student Symposium (GSS), Dauphin Island, AL

7. Perry GLW, Miller BP, Enright NJ (2006) A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. Plant Ecol 187:59-82

8. Williamson D, McLafferty S, Goldsmith V, McGuire P, Mollenkopf J (1998) Smoothing Crime Incident Data: New Methods for Determining the Bandwidth in Kernel Estimation

9. Williamson D, McLafferty S, Goldsmith V, Mollenkopf J, McGuire P (1999) A Better Method To Smooth Crime Incident Data ESRI ArcUser Magazine 1-5

10. Bailey TC, Gatrell AC (1995) Interactive Spatial Data Analysis. Longman Scientific and Technical, Essex, England

11. Fotheringham AS, Brunsdon C, Charlton M (2000) Quantitative Geography Perspectives on Spatial Data Analysis. SAGE Publications, London

12. Terrell GR (1990) The Maximal Smoothing Principle in Density Estimation. Journal of the American Statistical Association 85:470-477

13. Bowman A, Azzelini A (1997) Applied Smoothing Techniques for Data Analysis. Oxford University Press, Oxford

14. Lentz JA, Blackburn JK, Curtis AJ (2011) Evaluating Patterns of a White-Band Disease (WBD) Outbreak in Acropora palmata Using Spatial Analysis: A Comparison of Transect and Colony Clustering. PLoS one 6:e21830

15. Carr AP, Rodgers AR (2002) HRE: The Home Range Extension for ArcViewTM (Beta Test Version 0.9, July 1998). Tutorial Guide. Centre for Northern Forest Ecosystem Research, Ontario Ministry of Natural Resources, Ontario 22

16. Rodgers AR, Carr AP (2002) HRE: The Home Range Extension for ArcViewTM (Beta Test Version 0.9, July 1998). User's Manual. Centre for Northern Forest Ecosystem Research, Ontario Ministry of Natural Resources, Ontario 27

17. Rodgers AR, Kie JG (2010) HRT: Home Range Tools for ArcGIS(R) (Version 1.1, June 2007). User's Manual (Draft September 28, 2010). Centre for Northern Forest Ecosystem Research, Ontario Ministry of Natural Resources, Ontario 26

18. Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London

19. Worton BJ (1989) Kernel methods for estimating the utilization distribution in home-range studies. Ecology 70:164-168

20. Worton BJ (1995) Using Monte Carlo simulation to evaluate kernel-based home range estimators. J Wildl Manage 59:794-800

21. Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically weighted Regression: the analysis of spatially varying relationships. John Wiley & Sons, West Sussex

22. Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) Numerical recipes: the art of scientific computing. Cambridge Univ. Press, Cambridge, UK

23. Sain SR, Baggerly KA, Scott DW (1994) Cross-Validation of multivariate densities. Journal of the American Statistical Association 89:807-817

24. Jones MC, Marron JS, Sheather SJ (1996) A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association 91:401-407

25. Wand MP, Jones MC (1995) Kernel smoothing. Chapman and Hall, Ltd., London, UK

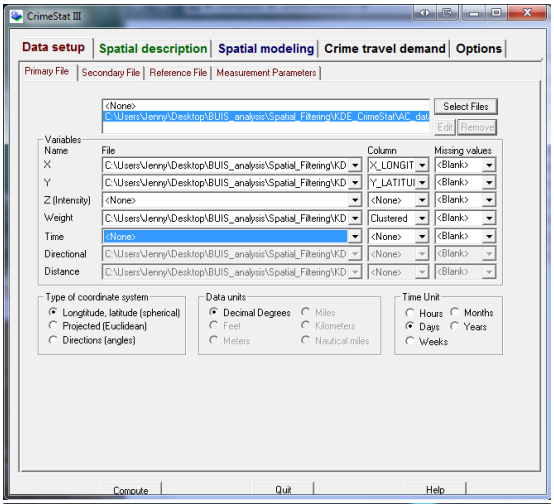
26. Williamson D (1999) K-NearestNeighbor (AS11177.zip). ArcScript., Available at: *http://arcscripts.esri.com/details.asp?dbid=11177*

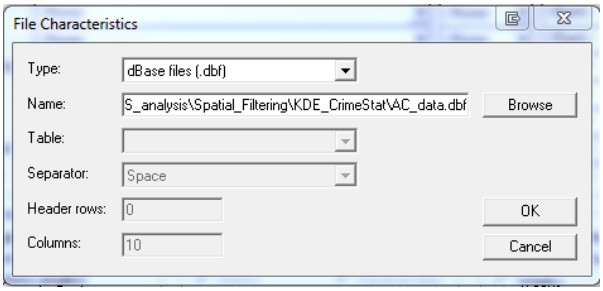
27. Levine N, Associates (2009) CrimeStat(R) III, version 3.2a: A Spatial Statistics Program for the Analysis of Crime Incident Locations. The National Institute of Justice, Washington, D.C. 1036

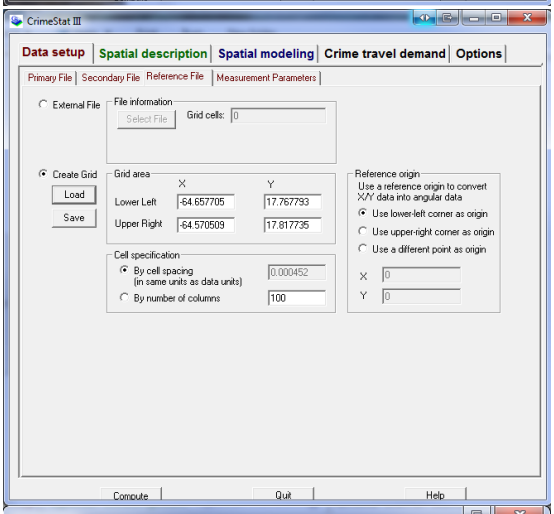
Appendix B – Analytical Procedures

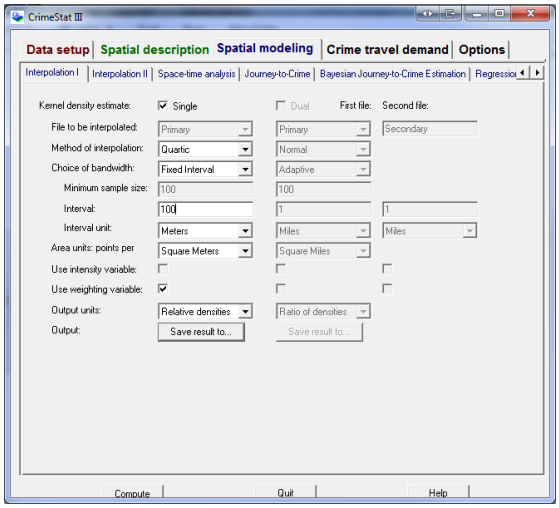
B.1 Kernel Density Estimates (KDE)

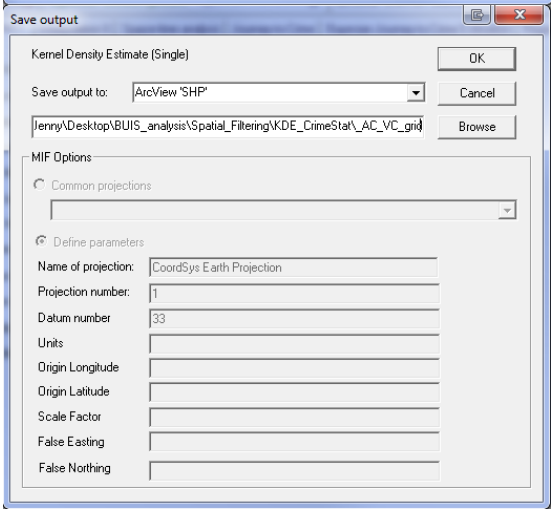
Step 1) Perform Single KDE analyses on the case data using ***CrimeStat III***

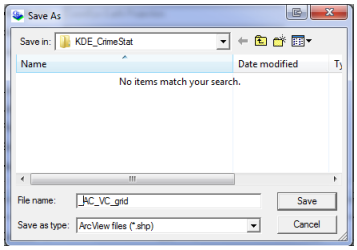


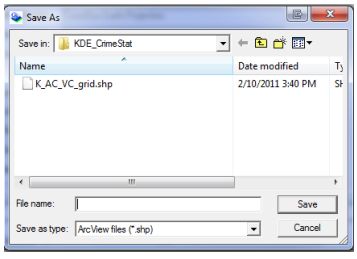




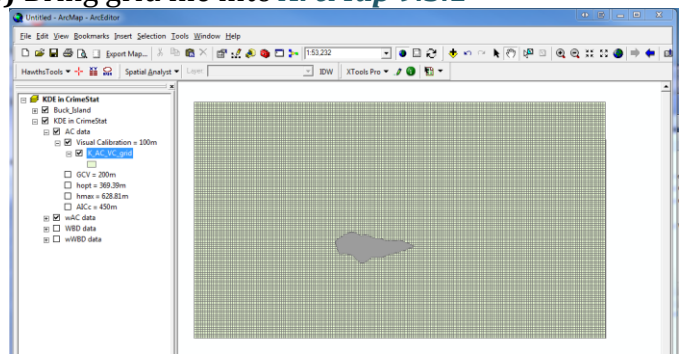






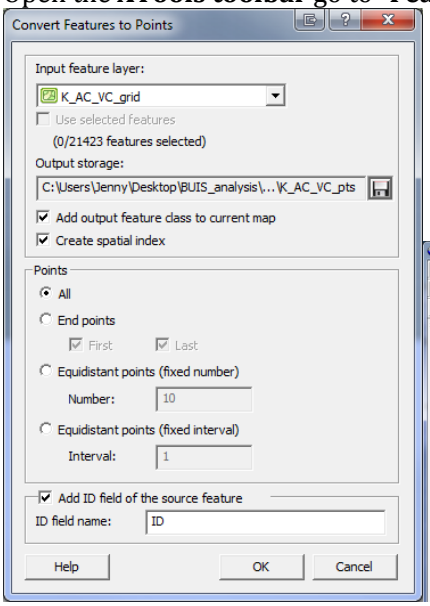


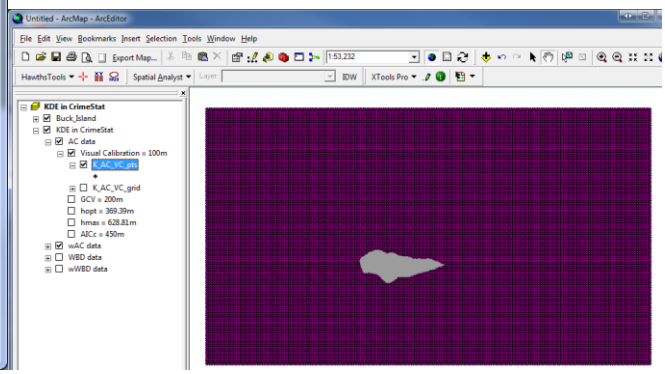
Step 2) Bring grid file into ***ArcMap 9.3.1***



Step 3) Download, Install, and then Enable the ***XTools Pro extension*** in ***ArcMap***

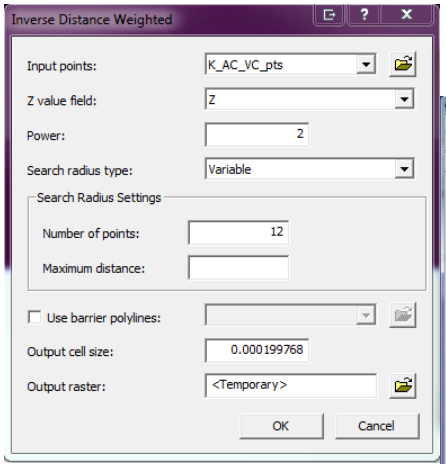
Open the **XTools toolbar** go to "**Feature Conversions**," select "**Convert Features to Points**"

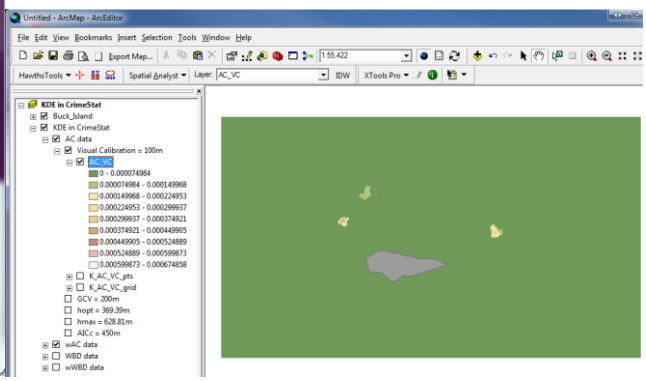




Step 4) Enable the ***Spatial Analyst Extension***

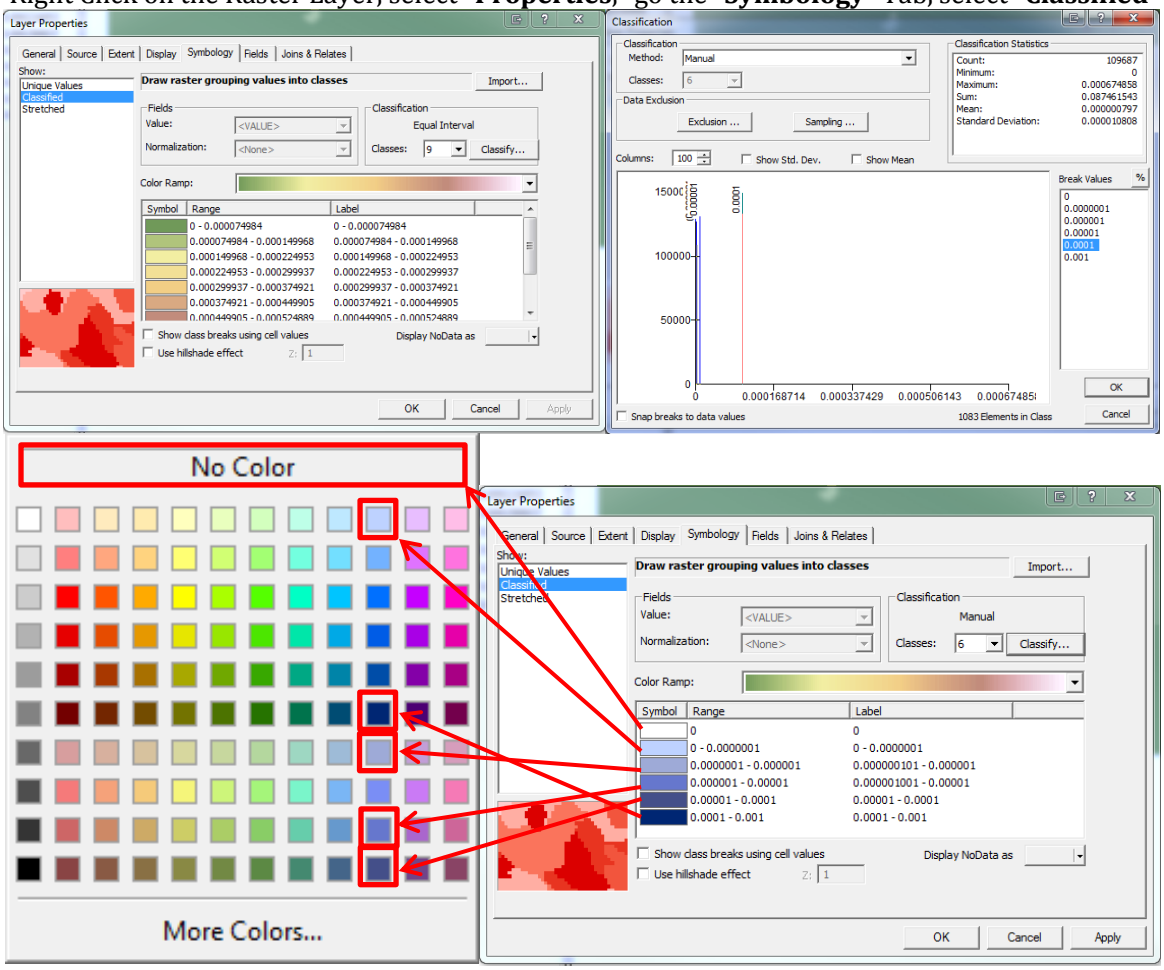
Open the Spatial Analyst toolbar, go to "**Interpolate to Rater**," select "**Inverse Distance Weighted**"





Step 5) Define the Colorscheme for the Newly Created Raster version of the KDE analysis

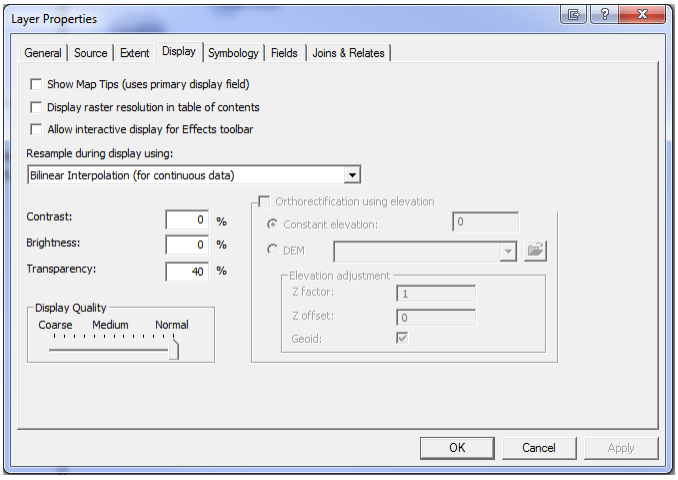
Right Click on the Raster Layer, select "**Properties**," go the "**Symbology**" Tab, select "**Classified**"



Step 6) Make the Raster Layer Semi-Transparent

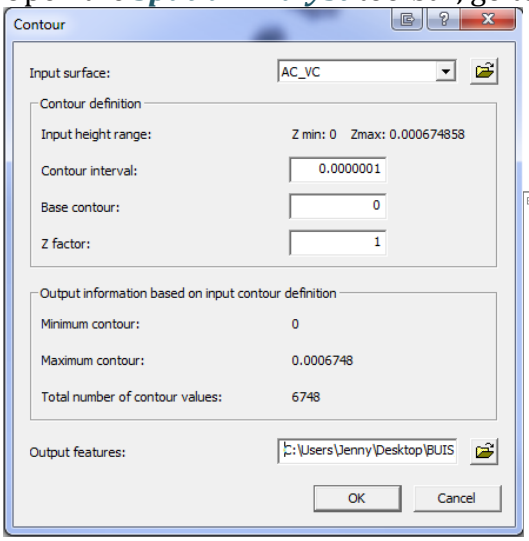
While still in the "**Properties**" window for the Raster Layer, go the "**Display**" Tab

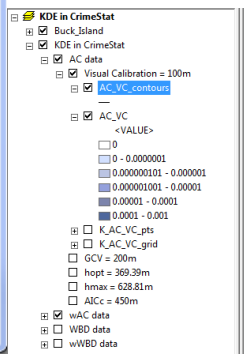
Set the **Transparency** to 40%



Step 7) Create Contours for the different KDE rates

Open the ***Spatial Analyst*** toolbar, go to "**Surface Analyst**" and select "**Contour…**"



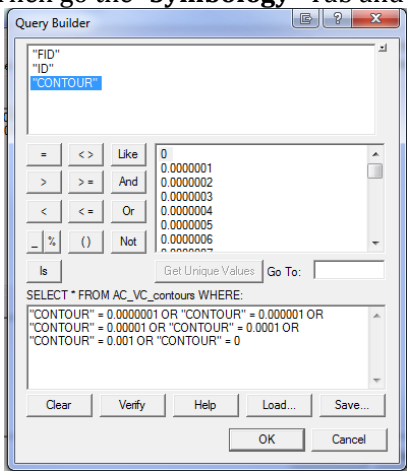


Step 8) Select Contours of Interest & then Format how they'll be displayed

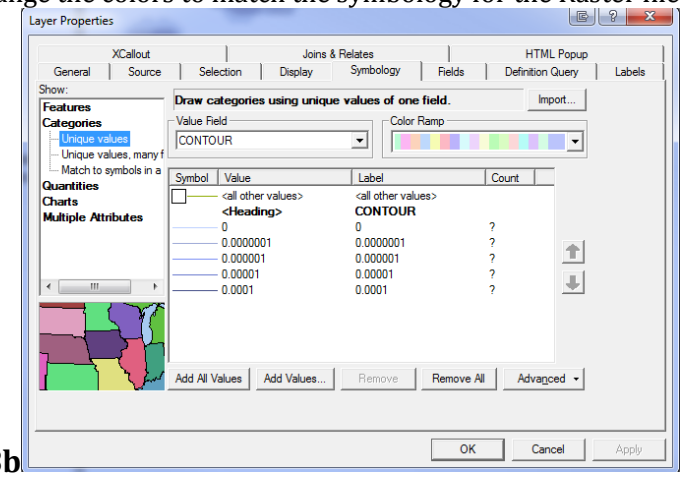
8a. Open the Contour Layer's "**Properties**," go the "**Definition Query**" Tab & select "**Query Builder**"

8b. Then go the "**Symbology**" Tab and change the colors to match the symbology for the Raster file

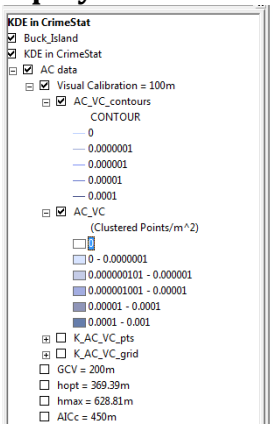
8a

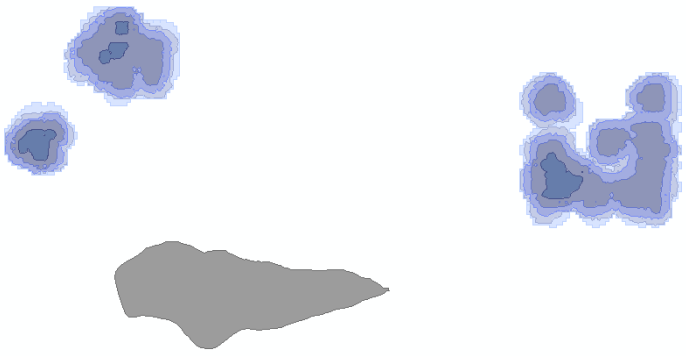


8b



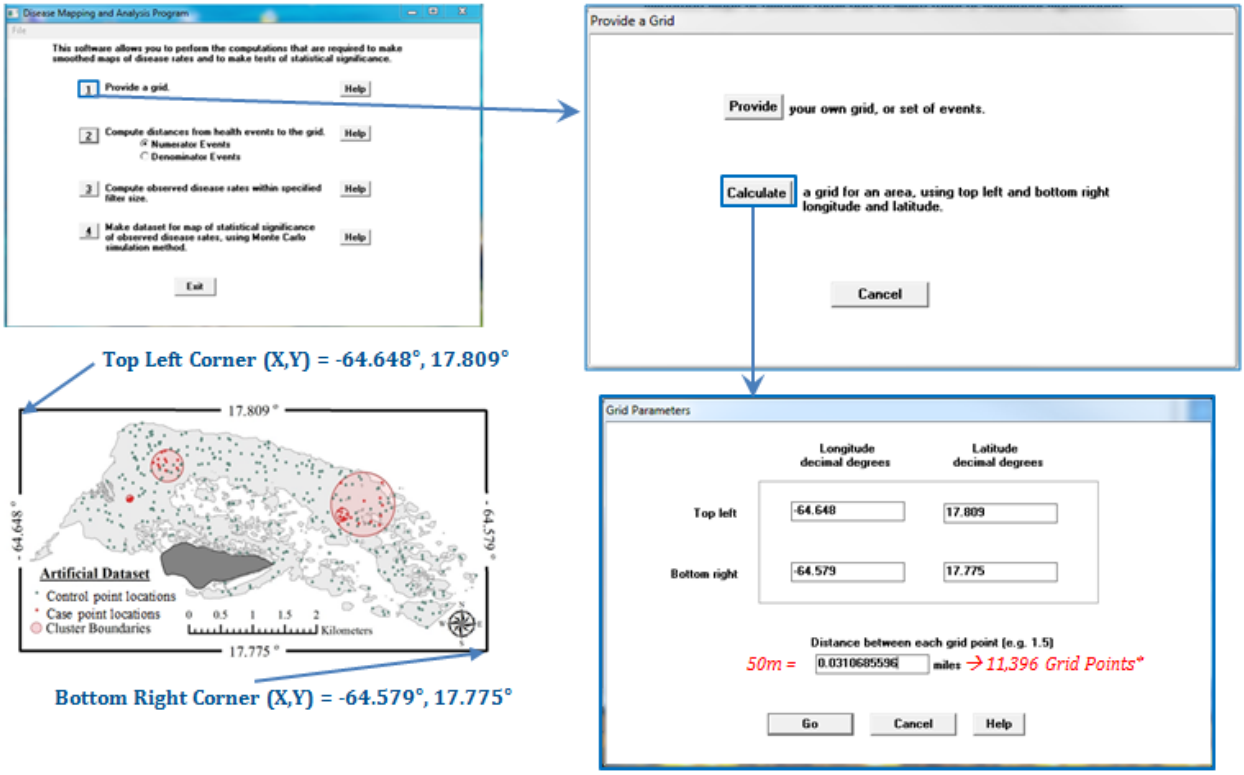
Step 9) Display the Finished KDE Product



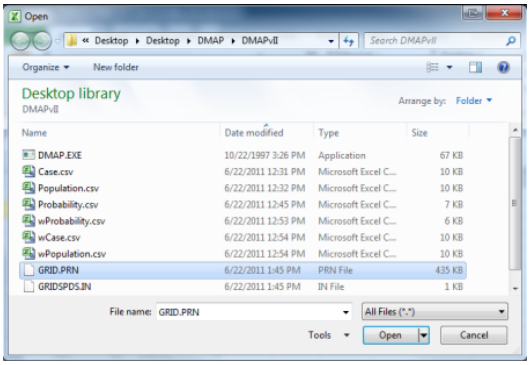


B.2 Spatial Filtering using Fixed Distance Filters in *DMAP*

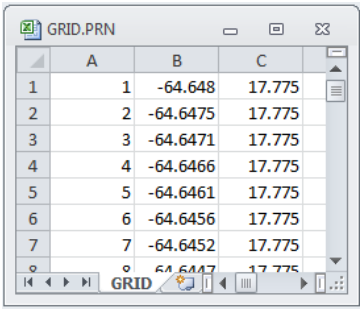
Step 1: Creating the Grid File



*DMAP II is a 16-bit program, this means it can only handle 2^{16} unique values, which is ±32,768 or 65,536 total values. Thus this version of DMAP can only handle datasets with <65,536 Total Grid Points, for this reason it's a good idea to check the total number of grid points in the newly generated "**GRID.PRN**"



To check this open the **GRID.PRN** file in Microsoft excel (it's usually easier to open the file from within Excel rather than double clicking on the actual file as your computer may not know that you want prn files to open in excel).



	A	B	C
1	1	-64.648	17.775
2	2	-64.6475	17.775
3	3	-64.6471	17.775
4	4	-64.6466	17.775
5	5	-64.6461	17.775
6	6	-64.6456	17.775
7	7	-64.6452	17.775

GRID.PRN

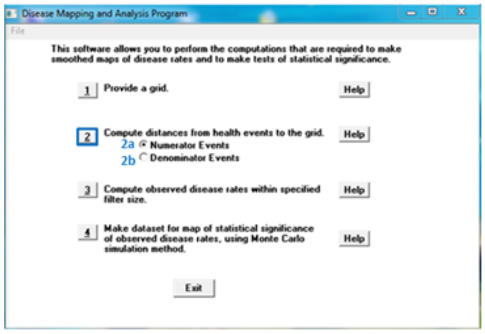
Column A = **Grid Point ID**

Column B = **Longitude** (in decimal degrees) for this location

Column C = **Latitude** (in decimal degrees) for this location

***Total # of Grid Points =** Count(A1:A11396) = **11,396 Total Grid Points**

Step 2: Computing the Distances from the health events to the Grid Points



We used Bowman & Azzelini's (1997) **Optimized Bandwidth (h_{opt}) statistic to calculate the best filter radius for this dataset:

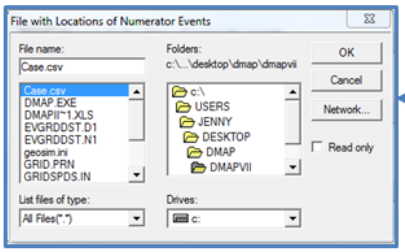
$$h_{opt} = \left\{\left[\frac{2}{(3)(n)}\right]^{\frac{1}{4}}\right\}\{\sigma\}$$

where n is the *sample size* calculated as the total # of point locations & σ is the *standard distance*

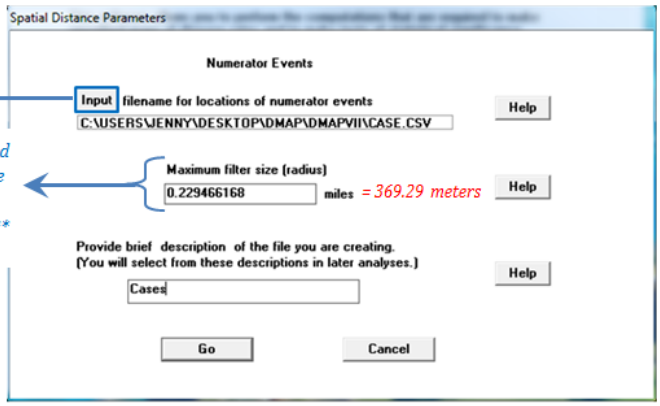
Thus, h_{opt} for both the weighted & non-weighted versions of our Artificial Cluster dataset would be calculated as follows:

$$h_{opt} = \left\{\left[\frac{2}{(3)(n)}\right]^{\frac{1}{4}}\right\}\{\sigma\} = \left\{\left[\frac{2}{(3)(375)}\right]^{\frac{1}{4}}\right\}\{1798.45m\} = 369.29m$$

2a: Computing the Distances from the Numerator (i.e. Case) Events to the Grid Points



Calculated using the h_{opt} statistic **

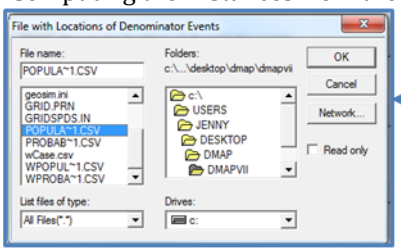


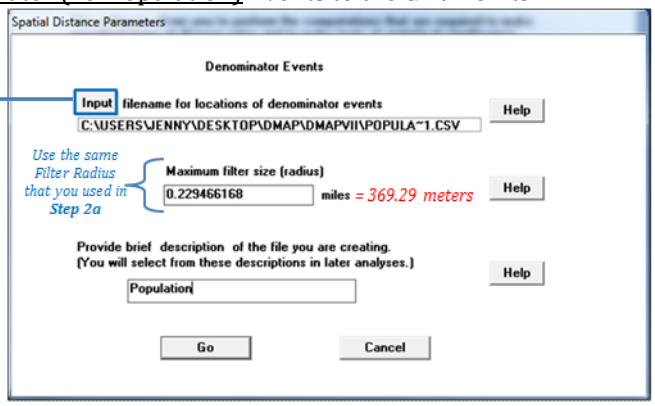
Start time= 4:30pm & End time= 4:42pm

Total Processing Time = 12 min

Resulting Numerator File = **EVGRDDST.N1**

2b: Computing the Distances from the Denominator (i.e. Population) Events to the Grid Points





Start time= 5:20pm & End time= 5:31pm

Total Processing Time = 11 min

Resulting Denominator File = **EVGRDDST.D1**

Steps 2a & 2b create the following 2 files: **EVGRDDST.N1** & **EVGRDDST.D1**

EVGRDDST.N1

	A	B	C	D
1	Cases			
18282	5432	56	0.1297	0
18283	5433	374	0.1943	1
18284	5433	371	0.211	1
18285	5433	368	0.1333	1
18286	5433	365	0.2077	1
18287	5433	353	0.2058	1
18288	5433	351	0.2105	1
18289	5433	349	0.2042	1
18290	5433	346	0.2237	1

EVGRDDST.D1

	A	B	C	D
1	Population			
18258	5431	371	0.1663	1
18259	5431	374	0.1885	1
18260	5432	56	0.1297	1
18261	5432	58	0.1413	1
18262	5432	91	0.1635	1
18263	5432	103	0.2087	1
18264	5432	115	0.0345	1
18265	5432	126	0.0584	1
18266	5432	161	0.2072	1

Cell A1 = Name you assigned to the numerator file, So for the Numerator file cell A1 = Cases, & for the Denominator file cell A1 = Population

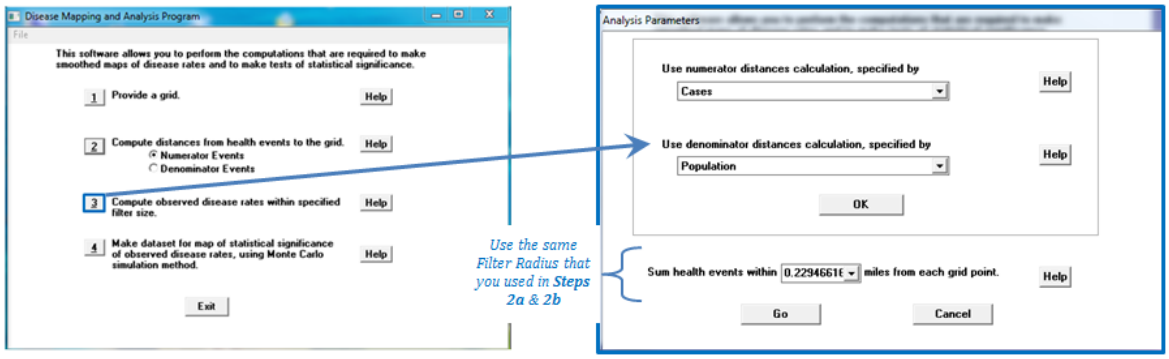
Column A = Grid Point IDs

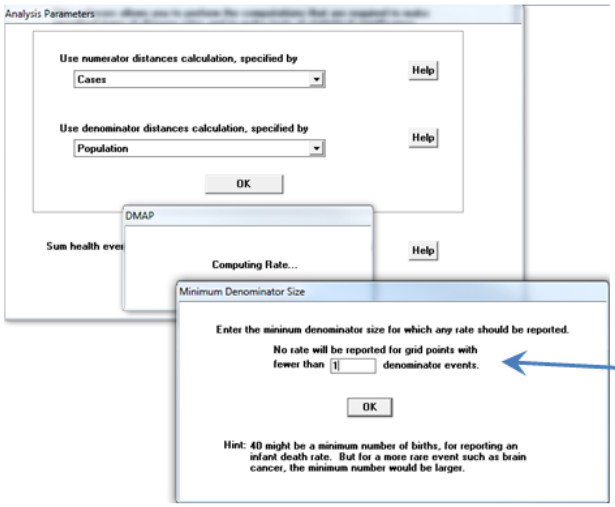
Column B = Data Point IDs

Column C = Distance (in miles) between the Grid Point & the Data Point

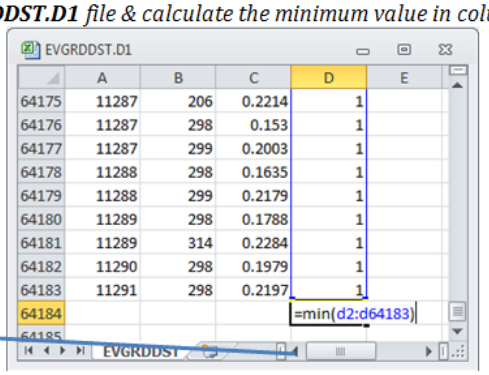
Column D = Weight of the Data Point

Step 3: Compute the Observed disease rates within specified filter size





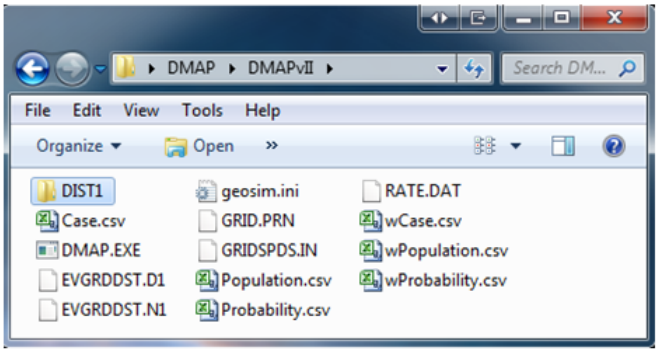
*To find the minimum number of denominator events open the **EVGRDDST.D1** file & calculate the minimum value in column D*

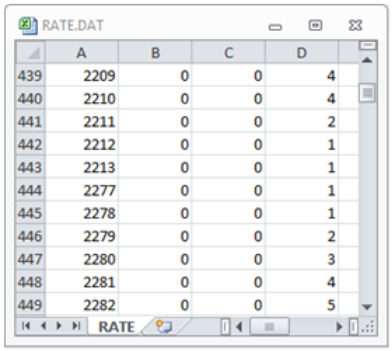


*For the **Non-Weighted** version of the data (**EVGRDDST.D1**) the minimum # of denominator events = 1*

Step 3 creates the following file
RATE.DAT

which is located in the newly created ***DIST1*** folder





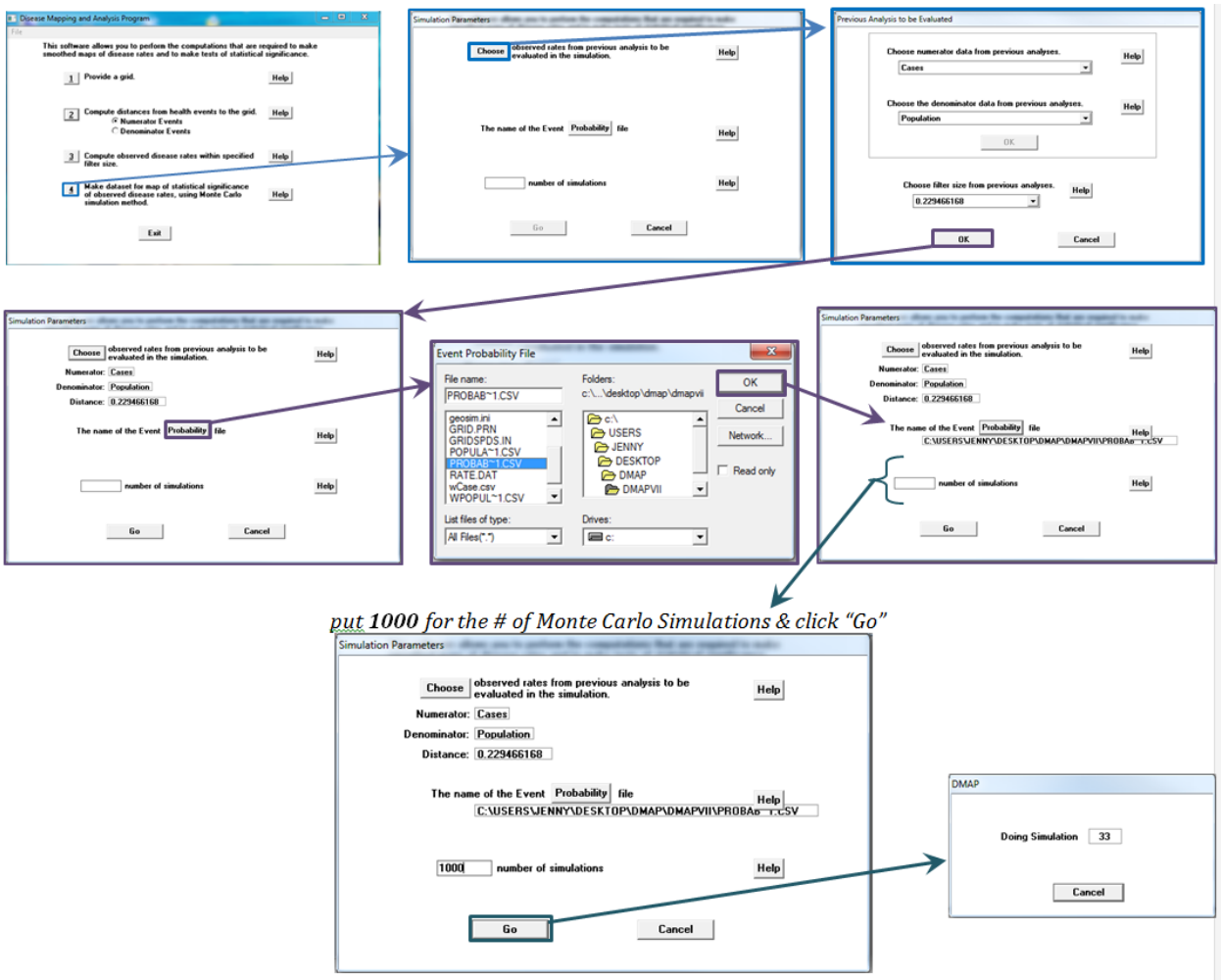
Column A = **Grid Point IDs**

Column B = Observed Event Rates within the specified filter size

Column C = Numerator Value associated with this grid point

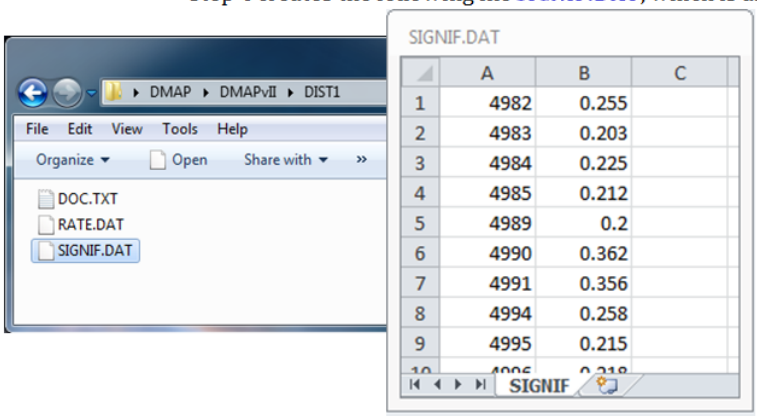
Column D = Denominator Value associated with this grid point

Step 4: Estimate the significance of observed disease rates, using Monte Carlo Simulations



*put **1000** for the # of Monte Carlo Simulations & click "Go"*

Step 4 creates the following file ***SIGNIF.DAT***, which is also located in the ***DIST1*** folder:



SIGNIF.DAT

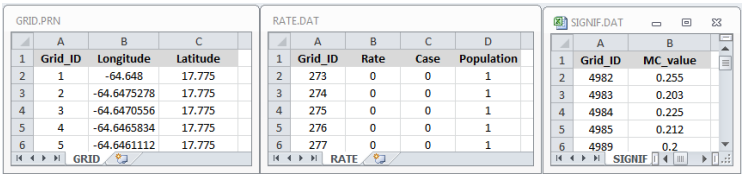
	A	B	C
1	4982	0.255	
2	4983	0.203	
3	4984	0.225	
4	4985	0.212	
5	4989	0.2	
6	4990	0.362	
7	4991	0.356	
8	4994	0.258	
9	4995	0.215	

Column A = Grid Point ID

Column B = Monte Carlo Simulated Value

Step 5: Import the GRID.PRN, RATE.DAT, & SIGNIF.DAT files into Excel

5a: Add a Header row to each of the 3 files



GRID.PRN

	A	B	C
1	Grid_ID	Longitude	Latitude
2	1	-64.648	17.775
3	2	-64.6475278	17.775
4	3	-64.6470556	17.775
5	4	-64.6465834	17.775
6	5	-64.6461112	17.775

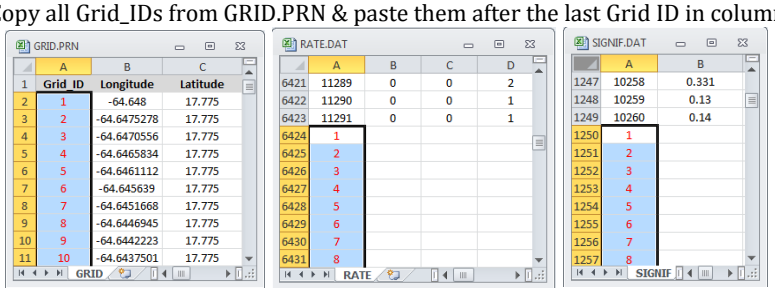
RATE.DAT

	A	B	C	D
1	Grid_ID	Rate	Case	Population
2	273	0	0	1
3	274	0	0	1
4	275	0	0	1
5	276	0	0	1
6	277	0	0	1

SIGNIF.DAT

	A	B
1	Grid_ID	MC_value
2	4982	0.255
3	4983	0.203
4	4984	0.225
5	4985	0.212
6	4989	0.2

5b: Copy all Grid_IDs from GRID.PRN & paste them after the last Grid ID in column A of both the RATE & SIGNIF files



GRID.PRN

	A	B	C
1	Grid_ID	Longitude	Latitude
2	1	-64.648	17.775
3	2	-64.6475278	17.775
4	3	-64.6470556	17.775
5	4	-64.6465834	17.775
6	5	-64.6461112	17.775
7	6	-64.645639	17.775
8	7	-64.6451668	17.775
9	8	-64.6446945	17.775
10	9	-64.6442223	17.775
11	10	-64.6437501	17.775

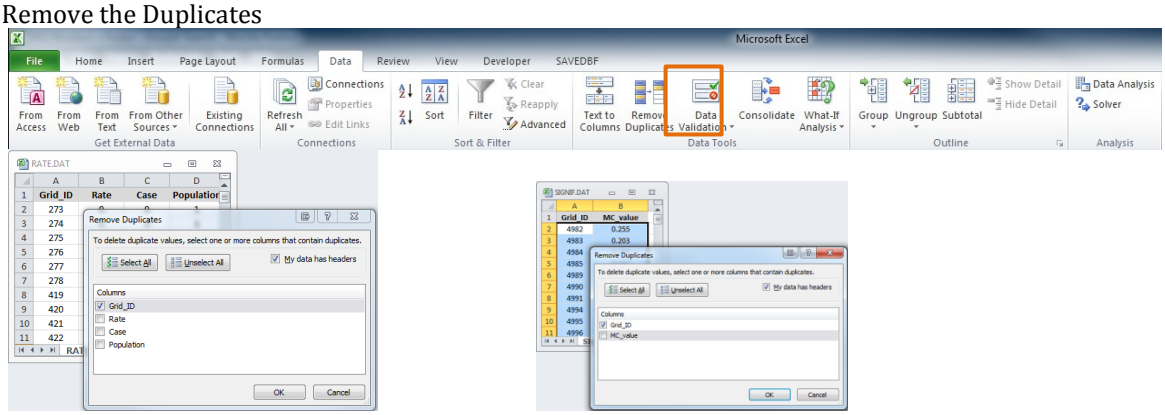
RATE.DAT

	A	B	C	D
6421	11289	0	0	2
6422	11290	0	0	1
6423	11291	0	0	1
6424	1			
6425	2			
6426	3			
6427	4			
6428	5			
6429	6			
6430	7			
6431	8			

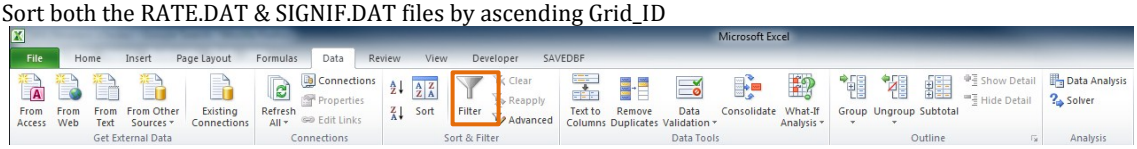
SIGNIF.DAT

	A	B
1247	10258	0.331
1248	10259	0.13
1249	10260	0.14
1250	1	
1251	2	
1252	3	
1253	4	
1254	5	
1255	6	
1256	7	
1257	8	

5c: Remove the Duplicates

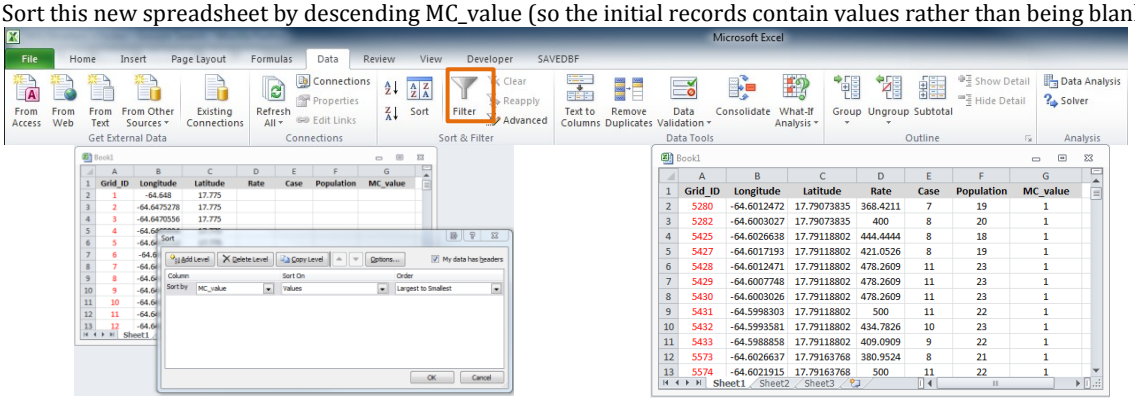


5d: Sort both the RATE.DAT & SIGNIF.DAT files by ascending Grid_ID



5e: Copy the Sorted Data from the 3 files into a new spreadsheet

5f: Sort this new spreadsheet by descending MC_value (so the initial records contain values rather than being blank)



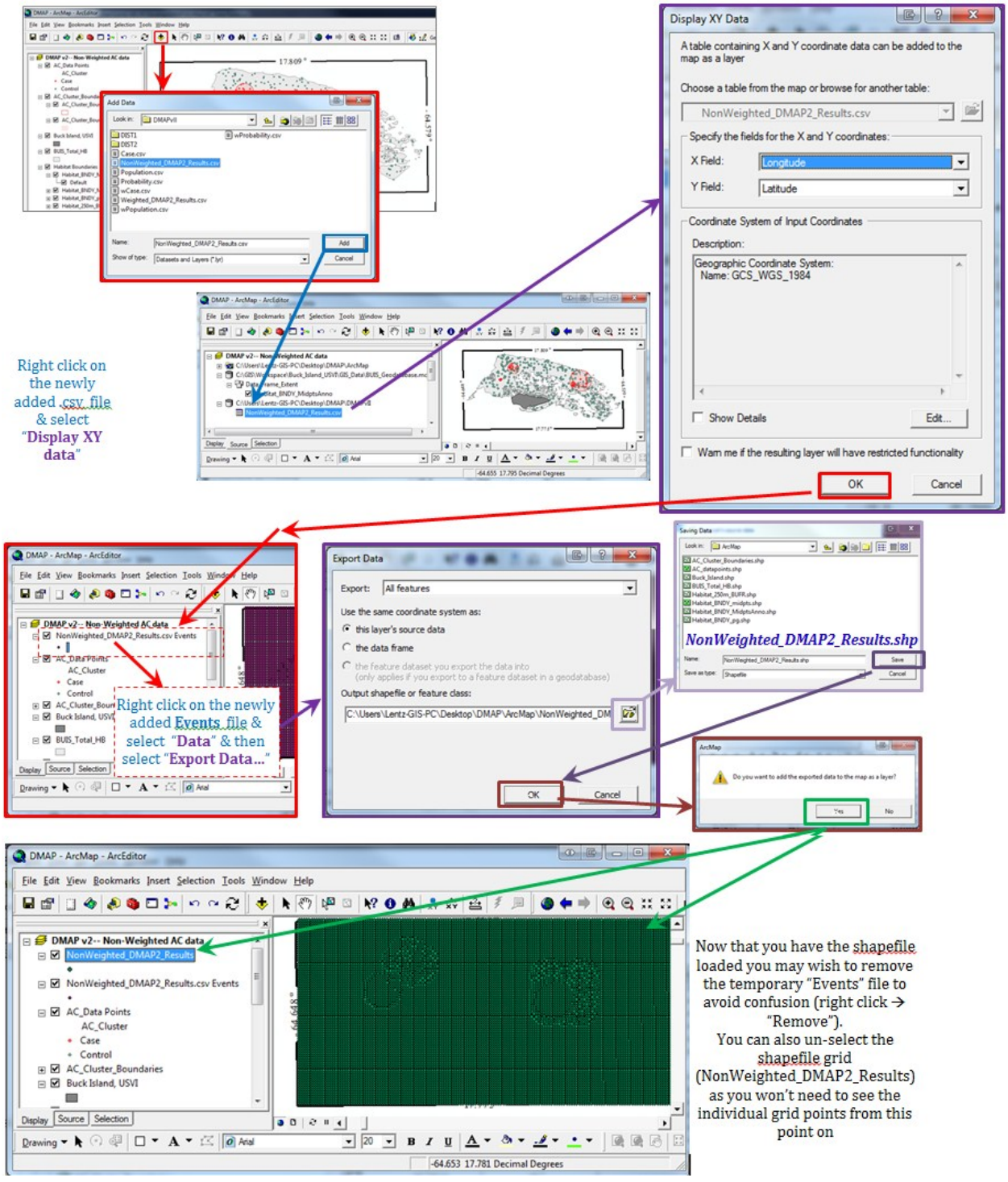
Book1

	A	B	C	D	E	F	G
1	Grid_ID	Longitude	Latitude	Rate	Case	Population	MC_value
2	5280	-64.6012472	17.79073835	368.4211	7	19	1
3	5282	-64.6003027	17.79073835	400	8	20	1
4	5425	-64.6026638	17.79118802	444.4444	8	18	1
5	5427	-64.6017193	17.79118802	421.0526	8	19	1
6	5428	-64.6012471	17.79118802	478.2609	11	23	1
7	5429	-64.6007748	17.79118802	478.2609	11	23	1
8	5430	-64.6003026	17.79118802	478.2609	11	23	1
9	5431	-64.5998303	17.79118802	500	11	22	1
10	5432	-64.5993581	17.79118802	434.7826	10	23	1
11	5433	-64.5988858	17.79118802	409.0909	9	22	1
12	5573	-64.6026637	17.79163768	380.9524	8	21	1
13	5574	-64.6021915	17.79163768	500	11	22	1

5g: Save this spreadsheet as either a DBFIV or a CSV file

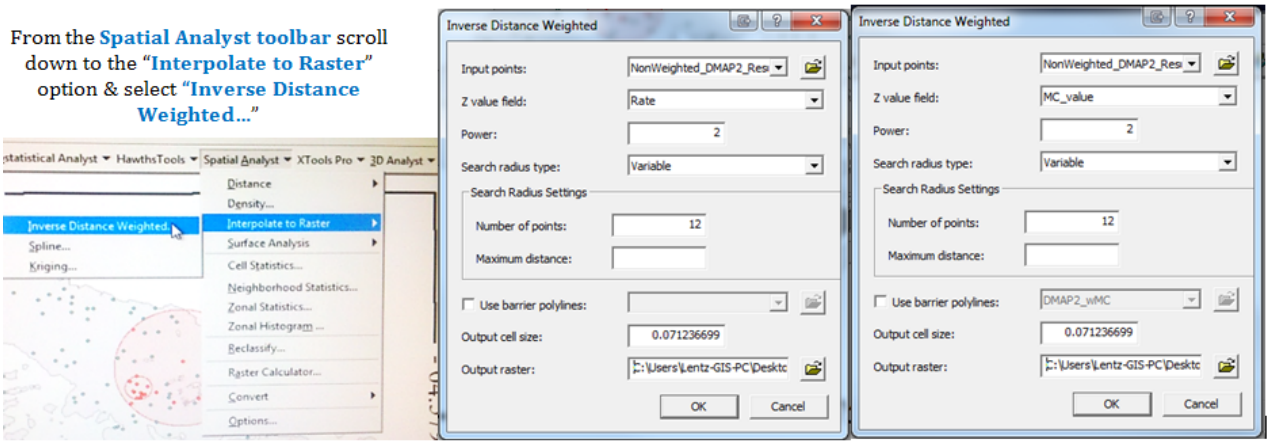
Step 6: Display *DMAP* results in *ArcMap*

6a: Import the NonWeighted_DMAP2_Results.csv file into *ArcMap*

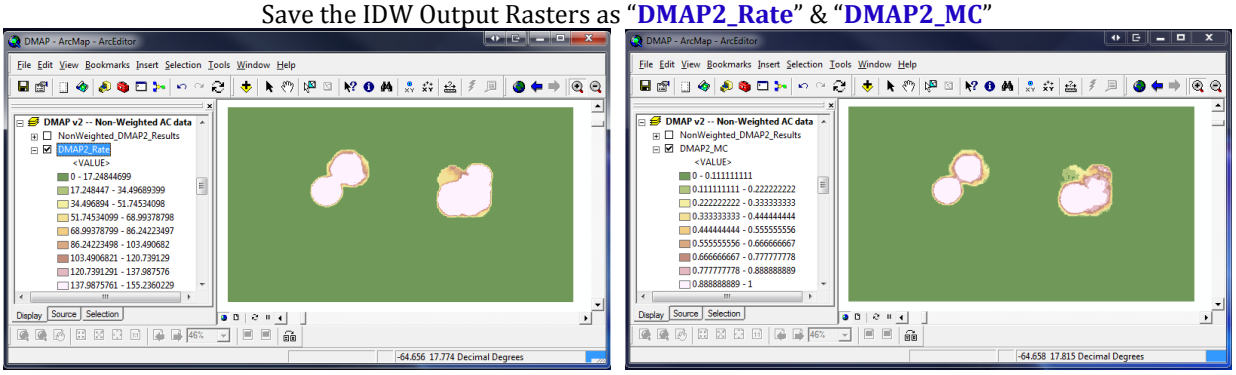


6b: Interpolate both the Clustering Rates & the Monte Carlo Significance values

From the **Spatial Analyst toolbar** scroll down to the "**Interpolate to Raster**" option & select "**Inverse Distance Weighted…**"



Save the IDW Output Rasters as "**DMAP2_Rate**" & "**DMAP2_MC**"



6c: Change the Symbology for the color-ramped Rate file as follows

Right Click on the "**DMAP2_Rate**" layer & select **Properties**, then click the **Symbology** Tab

Change the # of classes from 9 to **8** then click "**Classify**"

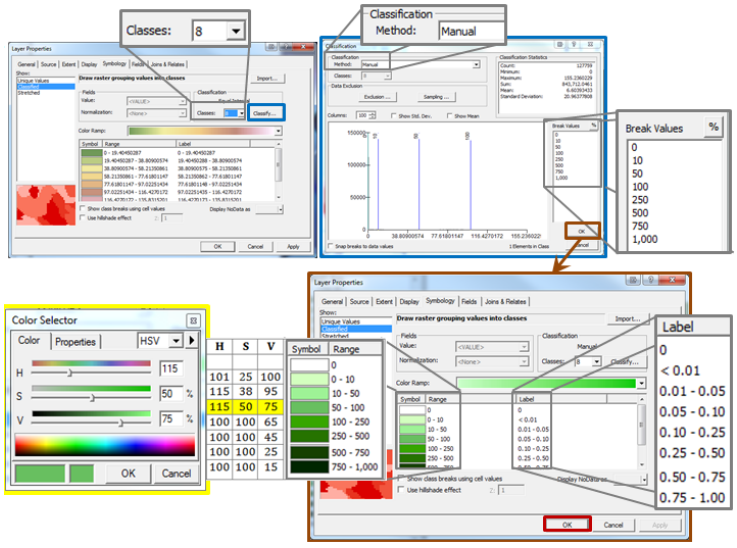
Change the **Classification Method** & **Break Values** so that they match the screen captures

Click **OK**

Now adjust the **Labels** & **Symbol colors** so that they match the screen captures on the right

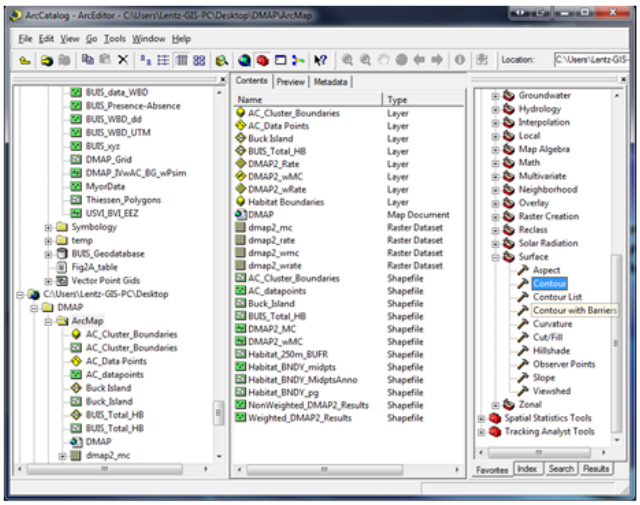
To change the colors right click on a color, click "**Properties for Selected Colors…**" then click "**More Colors…**" change the HSV values for each color so that they match the table to the right

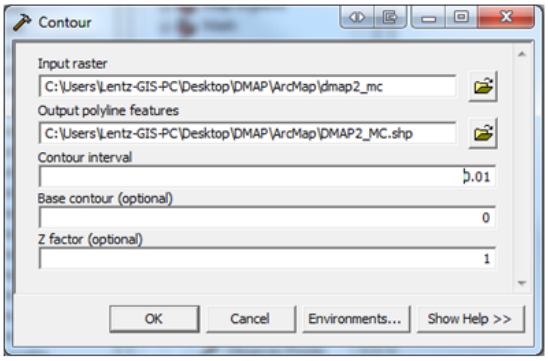
When you're satisfied with the revised symbology for the Clustering Rates color-ramped raster file, Click **OK**



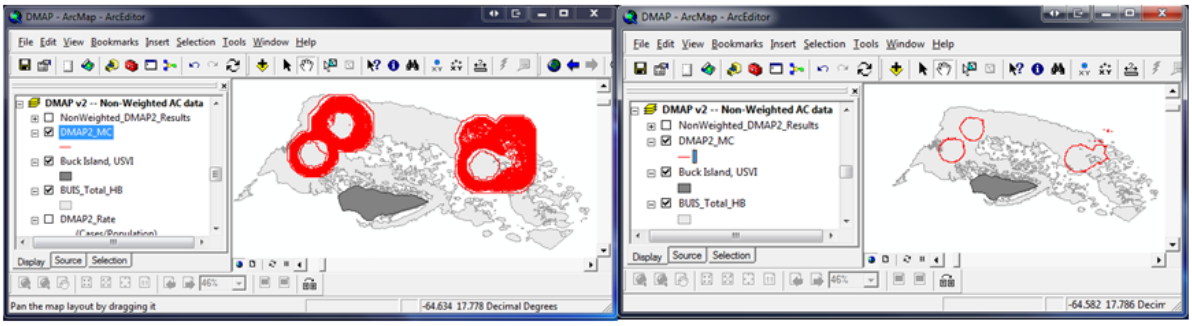
6d: Create Contours of the Monte Carlo Significance Raster data

Open **ArcCatalog** → open **ArcToolbox**→ expand the **Spatial Analyst Tools**→ expand the **Surface** toolbox & select the **Contour** tool



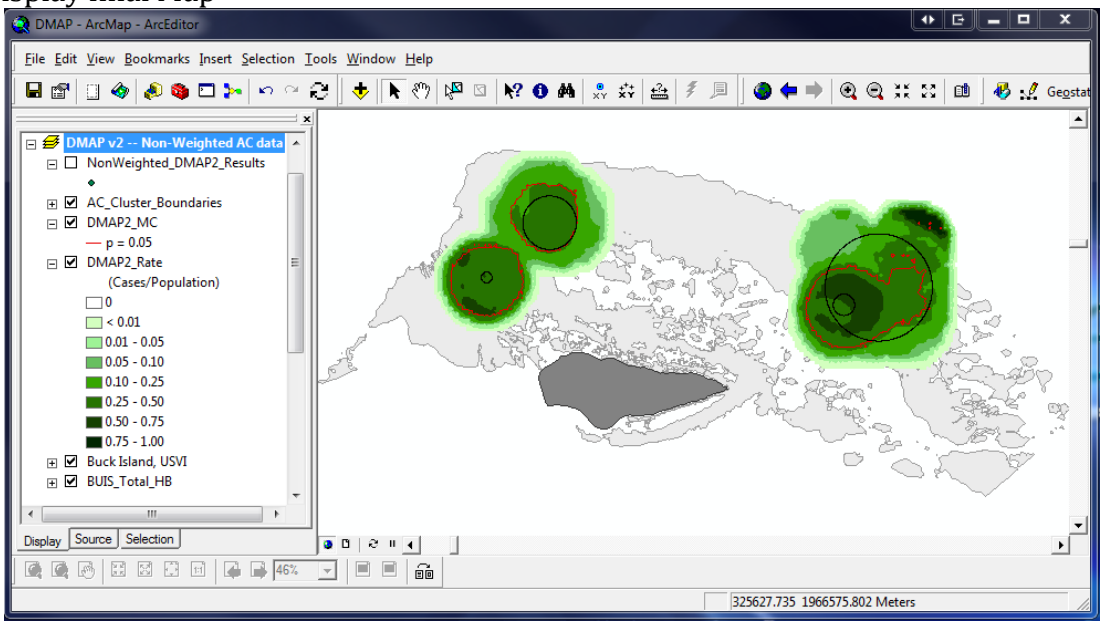


Return to ArcMap & add this new shapefile to your map (**DMAP2_MC.shp**)
Right click on this new layer, click **Properties**, select the **Definition Query** tab, & click **Query Builder** & type **"CONTOUR" = 0.94999999"** click **OK** & **OK** again



Now overlay these p = 0.05 contours over the clustering rates & compare results to the pre-defined cluster boundaries (shown in black)

6e: Display final Map



Appendix C – Supplemental Material from Chapter 8*

C.1 Supplemental Materials and Methods

The computed Ripley's *K* values were plotted against the distances tested in for each of the six coral types (**Figure C.1**). To facilitate interpretation of the above spatial distributions, I normalized the Ripley's *K* output by subtracting the expected values (*d*) from the observed values (*L(d)*), so that the new benchmark for evaluating complete spatial randomness (CSR) was y=0 (as opposed to the pre-normalized benchmark of πd^2). In order to determine whether or not clustering was present (and if so whether or not the aggregation was statistically significant) the normalized *K* values (*L(d) – d*) were then plotted against distance (**Figure C.2**). Next, to test the null hypothesis ($\boldsymbol{H_{C3}}$) that transect locations that are weighted by the number of colonies within them are not significantly more clustered (or dispersed) than the underlying pattern of just their locations, I plotted the weighted *K* using the confidence intervals (CIs) for the unweighted *K* (**Figure C.3**). I plotted the reverse combination (unweighted *K* and the CIs for the weighted *K*) in order to test the null hypothesis ($\boldsymbol{H_{C4}}$) that the colony-level dataset would be more clustered or dispersed than they would be by chance alone (**Figure C.4**).

C.2 Supplemental Results

The *Ripley's K* statistics were computed, the underlying coral population continued to show signs of significant clustering throughout all of the tested spatial distances at the transect-level (see **Figure C.1B** and **C.2B**), and for the most of the distances at the colony-level (see **Figure C.1D** and **C.2D**), with the highest degree of significant clustering occurring at distance

*Supplemental Material from **Chapter 8** has been reprinted with permission from *PLoS ONE* with slight modifications. For original publication please see Lentz JA, Blackburn JK, Curtis AJ (2011) Evaluating Patterns of a White-Band Disease (WBD) Outbreak in *Acropora palmata* Using Spatial Analysis: A Comparison of Transect and Colony Clustering. PLoS one 6:e21830

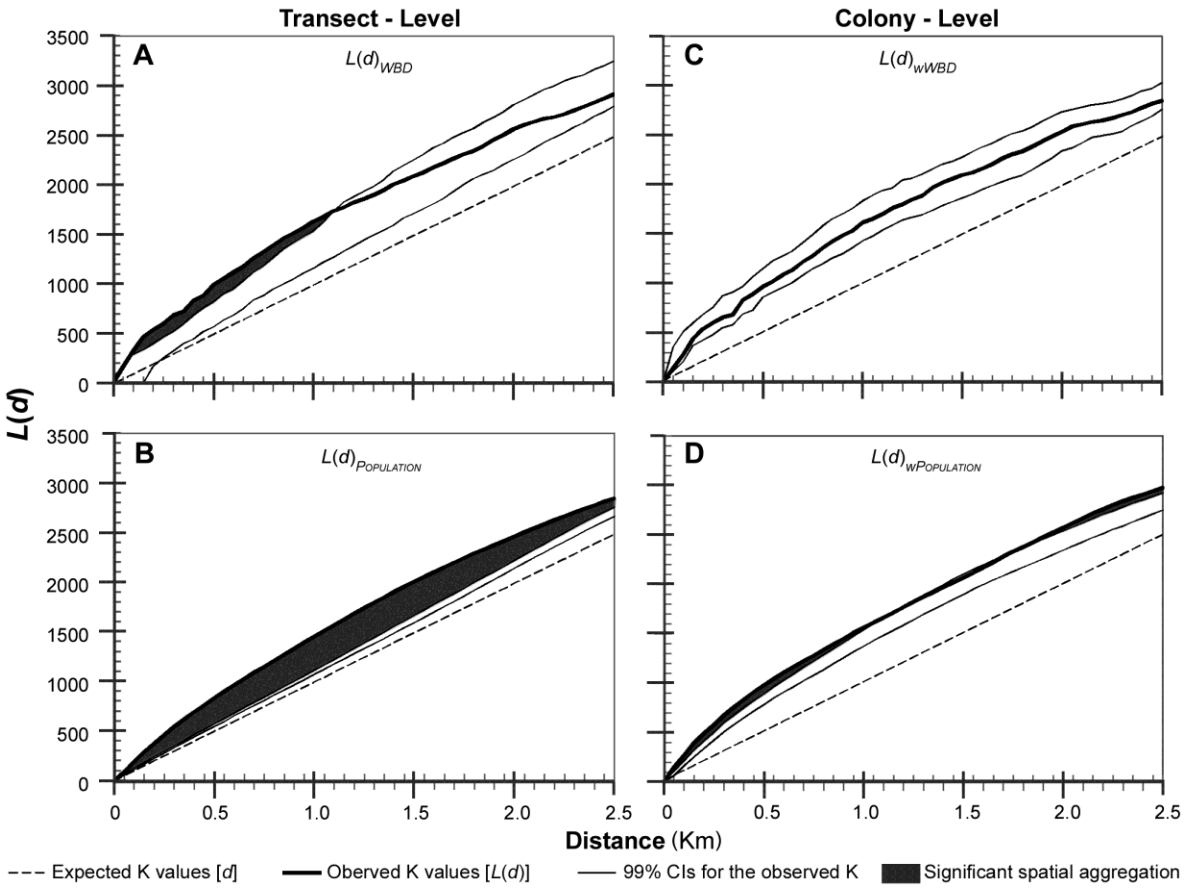


Figure C.1. *Ripley's K* plots of the diseased and underlying population at both the transect and colony-levels. *Ripley's K* plots comparing the spatial patterning of white-band disease (WBD) and the underlying *Acropora palmata* population, and showing the affect distance has on each of these spatial patterns. The null distribution of complete spatial randomness (CSR) is represented by the *Expected K* values (*d*) which are equal to the distance interval in which they are being tests (for example, the *Expected K* value at a distance of 500*m* would be 500), thus as the distance threshold increases so will the *Expected K* values. In all cases the *Observed K* (thick lines), and their corresponding 99% confidence intervals (thin lines) fell above the CSR benchmark (dashed line) indicating that both WBD and the underlying coral population had aggregated (clustered) spatial distributions across all of the tested distances at both the transect and colony-level. The results of the non-weighted *K* functions (**A–B**) assess the degree of clustering or dispersion present in the spatial distribution of the transect locations; while the results of the weighted *K* functions (**C–D**), in which each transect location was weighted by the number of colonies within it, evaluate the degree of clustering or dispersion of the colonies. (**A**) Significant clustering (shaded region) was detected in the spatial distribution of transects with WBD present at distances to ≤1.1*km*, and non-significant clustering was detected up to 2.5*km* (the maximum distance tested). (**B**) The spatial distribution of the 375 transects containing *A. palmata* showed significant clustering at all of the tested distances. (**C**) When the locations of transects with WBD present were weighted by the number of WBD colonies within them, their resulting spatial distribution was clustered, but not to a statistically significant extent. (**D**) When the transect locations of the underlying population were weighted by the total number of colonies within them, their resulting spatial distribution showed signs of aggregation at all of the distances tested, but only detected significant clustering at distances ≤1.05*km* and ≥1.75*km*.

thresholds of 1450*m* and 1700*m* for the transect- and colony-level datasets respectively (see **Figure C.2B**,**D**). Clustering was detected in the distribution of diseased corals for all of the tested distances, with the greatest degree of clustering occurring at distances of 1100*m* for the transect-level data (see **Figure C.2A**) and 1400*m* for the colony-level data (see **Figure C.2C**), with distances ≤ 1100*m* showing statistically significant spatial aggregation of WBD at the transect-level. In both versions of the dataset, the normalized *Observed K* for WBD takes a sharp dip at distance threshold of 350*m* (see circled regions of **Figure C.2A,C**). Overall, the plots based on the normalized *Ripley's K* values were preferred over plots based on the raw K values, because when the data were normalized such that the CSR benchmark was set to y=0, the hyperbolic nature of the plots was removed and the resulting graphs were much more expressive. **Figure C.3**, shows the graphical test of the null hypothesis ($\boldsymbol{H_{C3}}$) that transects weighted by the number of colonies within them are not significantly more clustered or dispersed than the underlying spatial distribution based on the transect locations alone. The $\boldsymbol{H_{C3}}$ hypothesis was rejected for WBD at distances < 1100*m* because the colony-level *Observed K* were greater than the upper CI for the transect-level *Observed K* indicating that transects weighted by the number of WBD colonies within them were significantly more clustered than their locations alone would suggest (see **Figure C.3A**). However, $\boldsymbol{H_{C3}}$ was accepted when WBD was examined at distances >1100*m*, as the *Observed K* for WBD colonies was within the upper and lower CI for the *Observed K* of the transects containing WBD, indicating that the spatial aggregation of WBD was not statistically significant at these distance scales. This hypothesis was rejected for the underlying population for all of the distance scales tested because the colony-level *Observed K* was above the upper CI for the transect-level population data, indicating that the transects weighted by the number of colonies within them were, in fact, significantly more clustered than the spatial distribution of the transect locations alone (see **Figure C.3B**).

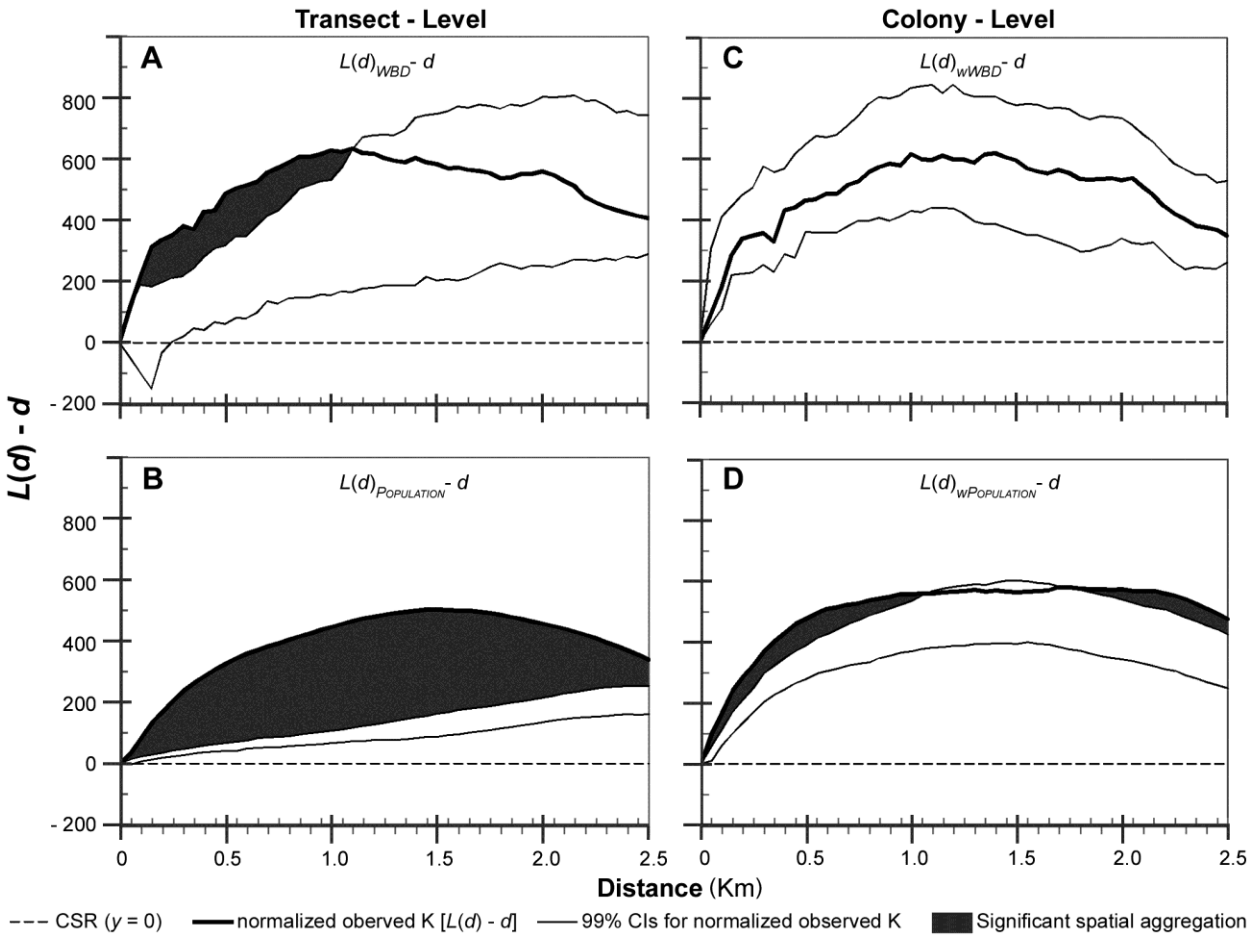


Figure C.2. Normalized *Ripley's K* plots depicting the same information as shown in **Figure C.1**. The transect locations for both white-band disease (WBD, **A**) and the underlying population (**B**) were clustered at all spatial distances tested (0–2.5*km*); with the population showing significant clustering (shaded region) at all distances <2.5*km* and significant clustering only occurring at distances ≤1.1*km* for transects in which WBD was present. (**C**) Transects containing WBD colonies still appear to be spatially aggregated across all of the tested spatial scales, but not to a statistically significant extent. (**D**) As in the transect-level analysis, the distribution of transects containing both diseased and non-diseased *A. palmata* colonies was also spatially aggregated; however, when the transects are weighted by the number of colonies within them, they only appear to have statistically significant clustering when tested using distances thresholds ≤1.15 or ≥1.7*km*.

Figure C.4 shows a graphical test of the null hypothesis ($\boldsymbol{H_{C4}}$) that transects weighted by the number of colonies within them would be more clustered or dispersed than they would be by chance alone. $\boldsymbol{H_{C4}}$ was rejected for both WBD (**Figure C.4A**) and the underlying population (see **Figure C.4B**) because the *Observed K* based on the transect-level data fell within the CI envelope based on the colony-level *Observed K*s.

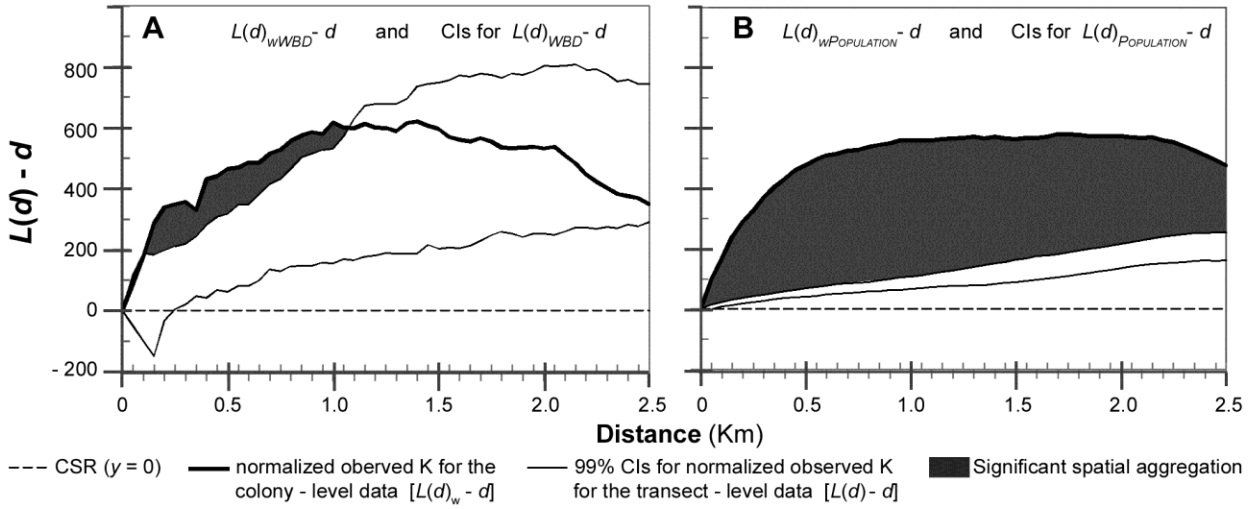


Figure C.3 Normalized *Ripley's K* Plots used to test the null hypothesis H_{C3}. Graphical representation of the test of the null hypothesis (H_{C3}) that transects weighted by the number of colonies within them will not be significantly more clustered or dispersed than the underlying spatial distribution based on the transect locations alone. In order for the null hypothesis to be accepted the *Observed K* based on the colony-level data (thick line) must fall within the upper and lower 99% confidence intervals (CIs, depicted as thin lines) estimated using the transect-level data. (**A**) The null hypothesis was rejected at distances <1.1*km* and accepted at distances >1.1*km* for white-band disease (WBD). (**B**) The null hypothesis was rejected for the population data at all of the distances tested.

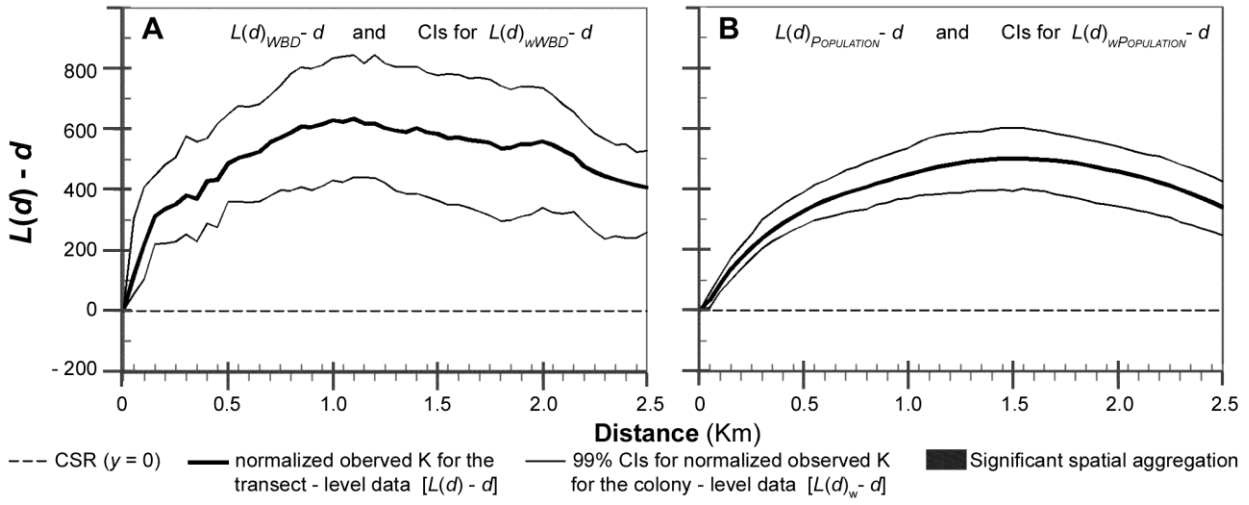


Figure C.4 Normalized *Ripley's K* Plots used to test the null hypothesis H_{C4}. A graphical representation of the test of the null hypothesis (H_{C4}) that the spatial distribution of the colony-level data would be more clustered or dispersed than they would be through chance alone. This hypothesis was rejected for both (**A**) white-band disease (WBD) and the (**B**) underlying population because the *Observed K* (thick line) based on the transect-level data falls within the 99% confidence intervals (CIs, depicted as thin lines) based on the *Observed K* estimated using the colony-level data.

Appendix D – Permission to Use Published Material in Dissertation

D.1 Letter of Release

Login | Create Account | Feedback

PLoS one

accelerating the publication of peer-reviewed science

Search articles... GO Advanced Search

Browse RSS

Home Browse Articles About For Readers For Authors and Reviewers Journals Hubs PLoS.org

Open-Access License

No Permission Required

The Public Library of Science (PLoS) applies the Creative Commons Attribution License (CCAL) to all works we publish (read the human-readable summary or the full license legal code). Under the CCAL, authors retain ownership of the copyright for their article, but authors allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in PLoS journals, so long as the original authors and source are cited. **No permission is required from the authors or the publishers.**

In most cases, appropriate attribution can be provided by simply citing the original article (e.g., Kaltenbach LS et al. (2007) Huntingtin Interacting Proteins Are Genetic Modifiers of Neurodegeneration. *PLoS Genet* 3(5): e82. doi:10.1371/journal.pgen.0030082). If the item you plan to reuse is not part of a published article (e.g., a featured issue image), then please indicate the originator of the work, and the volume, issue, and date of the journal in which the item appeared. For any reuse or redistribution of a work, you must also make clear the license terms under which the work was published.

This broad license was developed to facilitate open access to, and free use of, original works of all types. Applying this standard license to your own work will ensure your right to make your work freely and openly available. Learn more about open access. For queries about the license, please contact us.

All site content, except where otherwise noted, is licensed under a Creative Commons Attribution License.

Privacy Statement Terms of Use Advertise Media Inquiries PLoS in Print Site Map PLoS.org

Ambra 1.7 Managed Colocation provided by Internet Systems Consortium.

D.2 First Page from Publication

OPEN ACCESS Freely available online

PLoS one

Evaluating Patterns of a White-Band Disease (WBD) Outbreak in *Acropora palmata* Using Spatial Analysis: A Comparison of Transect and Colony Clustering

Jennifer A. Lentz[1]*, Jason K. Blackburn[2], Andrew J. Curtis[3]

1 Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana, United States of America, **2** Emerging Pathogens Institute and the Department of Geography, University of Florida, Gainesville, Florida, United States of America, **3** Department of American Studies and Ethnicity, University of Southern California, Los Angeles, California, United States of America

Abstract

Background: Despite being one of the first documented, there is little known of the causative agent or environmental stressors that promote white-band disease (WBD), a major disease of Caribbean *Acropora palmata*. Likewise, there is little known about the spatiality of outbreaks. We examined the spatial patterns of WBD during a 2004 outbreak at Buck Island Reef National Monument in the US Virgin Islands.

Methodology/Principal Findings: Ripley's K statistic was used to measure spatial dependence of WBD across scales. Localized clusters of WBD were identified using the DMAP spatial filtering technique. Statistics were calculated for colony- (number of *A. palmata* colonies with and without WBD within each transect) and transect-level (presence/absence of WBD within transects) data to evaluate differences in spatial patterns at each resolution of coral sampling. The Ripley's K plots suggest WBD does cluster within the study area, and approached statistical significance ($p = 0.1$) at spatial scales of 1100 m or less. Comparisons of DMAP results suggest the transect-level overestimated the prevalence and spatial extent of the outbreak. In contrast, more realistic prevalence estimates and spatial patterns were found by weighting each transect by the number of individual *A. palmata* colonies with and without WBD.

Conclusions: As the search for causation continues, surveillance and proper documentation of the spatial patterns may inform etiology, and at the same time assist reef managers in allocating resources to tracking the disease. Our results indicate that the spatial scale of data collected can drastically affect the calculation of prevalence and spatial distribution of WBD outbreaks. Specifically, we illustrate that higher resolution sampling resulted in more realistic disease estimates. This should assist in selecting appropriate sampling designs for future outbreak investigations. The spatial techniques used here can be used to facilitate other coral disease studies, as well as, improve reef conservation and management.

Citation: Lentz JA, Blackburn JK, Curtis AJ (2011) Evaluating Patterns of a White-Band Disease (WBD) Outbreak in *Acropora palmata* Using Spatial Analysis: A Comparison of Transect and Colony Clustering. PLoS ONE 6(7): e21830. doi:10.1371/journal.pone.0021830

Editor: Christian R. Voolstra, King Abdullah University of Science and Technology, Saudi Arabia

Received February 3, 2011; **Accepted** June 12, 2011; **Published** July 19, 2011

Funding: The enclosed manuscript contains analyses performed as part of an ongoing doctoral research project. This study has no direct funding source. The lead author is partially funded through a graduate fellowship from Louisiana State University. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Jennifer.Lentz@gmail.com

Introduction

Over the past three decades, the incidence of coral disease has increased from sparse, localized sightings, to an apparent panzootic, as disease sightings have become commonplace among the world's reef systems. Since the first documented cases of coral disease in the late 1960s and early 1970s [1–4], scientists have been working to identify causes of these diseases [5,6]; however, progress has been slowed by the complexity of coral ecosystems and anthropogenic influences on these systems [5–15]. Given the corresponding increase in human population pressure during this time period, it has been suggested that anthropogenic related stressors are contributing to, if not directly causing, coral disease outbreaks [5,9,16–23]. While correlations between anthropogenic stressors and disease frequencies have been seen for quite some time [15,17,24–27], it was only recently that direct experimental evidence was able to actually show how anthropogenic stress factors (such as climate change, water pollution, and overfishing) were directly contributing to coral disease [6,26,28,29].

While coral diseases are occurring globally, their incidence appears to be the most severe in the Caribbean [9,11,12,26,30–39]. Over the past few decades reports show that disease is responsible for a roughly 80% loss in Caribbean coral cover [24,40,41]. Within the Caribbean, the *Acropora* coral genus appears to have been the hardest hit by disease, with *A. palmata* showing a 90–95% decline [12,42–44] and *A. cervicornis* populations collapsing across the region [41,42,45,46], causing them to be the first corals in history to be listed as "threatened" under the United States Endangered Species Act.

In 1977, shortly after the first documented coral disease, black-band disease (BBD) [1,2], a second "band" disease was also discovered in the Caribbean [3,44]. This new white-band disease

Vita

Jennifer Anne Lentz was born in 1982 in Tucson, Arizona. She attended St. Gregory College Preparatory School, where she received the Studio Art Award and Science Research Award, graduating with High Honors in 2001.

She attended Hamilton College in Clinton, New York, where she designed her own major in environmental science with a marine concentration. Jennifer received the "Alexander Hamilton 100 Best Writers Award" in 2002 and 2003. In 2004, she spent her Spring semester studying at the School for Field Studies "Marine Resource" program located in the Turks and Caicos Islands. It was during this semester abroad that she first became aware of the declining health of Caribbean coral reefs, prompting her to do her semester long research project studying diseased *Acropora palmata* coral populations. Upon returning from the Turks and Caicos, Jennifer received a fellowship from Jug Bay Wetlands Sanctuary, in Lothian, Maryland, where she spent the summer using telemetry to track Eastern Box turtles. The director of the sanctuary was so impressed by her work, that he agreed to let her analyze the sanctuary's 9-year dataset as part of her undergraduate thesis, which used spatial analysis to study the home range and habitat preferences of Eastern Box turtles. Shortly before graduating Jennifer was invited to join Sigma Xi, a scientific honors society. In May 2005, Jennifer earned a Bachelor of Arts degree from Hamilton College, graduating *Cum Laude* with an Interdisciplinary Concentration in Environmental Science and Studio Art minor. Upon graduating, she also received Hamilton's Interdisciplinary Concentration Honor, and was awarded the Elihu Root Fellowship, which is given to students who have demonstrated high achievement and

special aptitude for scientific research and are planning to pursue graduate degree in the sciences.

The following fall, Jennifer began her doctoral studies as a Board of Regents Graduate Fellow in the Department of Oceanography and Coastal Sciences at Louisiana State University (LSU). Her graduate research combines her passion for coral reefs with her interest in geospatial analysis, with her dissertation focusing on developing analytical protocols for studying the spatial nature of coral diseases. In 2010, she was elected president of the Coast and Environment Graduate Organization (CEGO). She served as president for one year, during which time she received LSU's Tom W. Dutton Service Award, a Certificate of Appreciation from the German Embassy for CEGO's contribution to the Transatlantic Climate Bridge initiative, and the CEGO Leadership and Service Award from LSU's School of the Coast and Environment. Her doctoral work was also funded by a Graduate Research Fellowship and Teaching Assistantships through the Department of Oceanography and Coastal Sciences.

In May 2012, Jennifer will receive her Doctor of Philosophy degree from the Department of Oceanography and Coastal Sciences with a minor in geography.